

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:06:21 ; Search time 103.196 Seconds  
(without alignments)  
5032.400 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVTAHN.....NPQVASALTDLKEGLEMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

1: \_geneseq1980s.\*

2: \_geneseq1990s.\*

3: \_geneseq2000s.\*

4: \_geneseq2001s.\*

5: \_geneseq2002s.\*

6: \_geneseq2003as.\*

7: \_geneseq2003bs.\*

8: \_geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 AAW98011	AAW98011 Hypersens
2	9448	100.0	1838	3 AAY71095	Aay71095 Erwinia a
3	9448	100.0	1838	3 AAY84856	Aay84856 A hyperse
4	9448	100.0	1838	4 AAU02878	AAU02878 Erwinia a
5	9448	100.0	1838	5 AAEL6449	AAEL6449 Erwinia a
6	2723	28.8	518	4 AAU04490	AAU04490 Erwinia a
7	2563	27.1	507	4 AAU04491	AAU04491 Erwinia a
8	1041	11.0	201	3 AAY69258	Aay69258 Amino aci
9	302	3.2	3041	6 ABM70325	ABM70325 Phototrab
10	295.5	3.1	2975	6 ABM68759	ABM68759 Phototrab
11	295	3.1	2951	6 ABM67717	ABM67717 Phototrab
12	280.5	3.0	4560	6 ABM67454	ABM67454 Phototrab
13	275	2.9	2514	3 AAY75097	Aay75097 Neisseria
14	275	2.9	2514	6 ABU06028	ABU06028 N. mening
15	275	2.9	2514	6 ABU38197	ABU38197 Protein e
16	274.5	2.9	9535	6 ABM73008	ABM73008 Staphyloc
17	274.5	2.9	10498	6 ABU19119	ABU19119 Pathogen
18	273.5	2.9	6713	6 ABU15887	ABU15887 Protein e
19	267.5	2.8	2659	5 AAY75489	Aay75489 S. aureus
20	262.5	2.8	5024	4 AAG82935	Aag82935 S. epider
21	262	2.8	6281	4 AAU37403	AAU37403 Staphyloc
22	261	2.8	2799	6 ABU37640	ABU37640 Protein e
23	258.5	2.7	6641	6 ABU42656	ABU42656 Protein e
24	257.5	2.7	10182	5 ABP38314	ABP38314 Staphyloc
25	257	2.7	5795	4 AAU37017	AAU37017 Staphyloc

26	256.5	2.7	2398	6 ABU42252	ABU42252 Protein e
27	253	2.7	2045	6 ABU46539	ABU46539 Protein e
28	253	2.7	2059	5 ABP25711	ABP25711 Streptoco
29	248.5	2.6	3158	4 AAU37018	AAU37018 Staphyloc
30	247.5	2.6	2344	4 AAU37120	AAU37120 Staphyloc
31	245	2.6	2599	3 AAY75098	Aay75098 Neisseria
32	245	2.6	2703	6 ABU38184	ABU38184 Protein e
33	244.5	2.6	3259	7 ADE56037	ADE56037 Human pro
34	244.5	2.6	3259	7 ADE56033	ADE56033 Staphyloc
35	242	2.6	2434	4 AAU34339	AAU34339 Protein e
36	241.5	2.6	1837	3 AAY85564	Aay85564 Human hom
37	241	2.6	1577	6 ABU41145	ABU41145 Protein e
38	240.5	2.5	2086	4 AAU34143	AAU34143 Staphyloc
39	240.5	2.5	5533	4 ABB65772	ABB65772 Drosophil
40	240.5	2.5	5560	4 ABB71160	ABB71160 Drosophil
41	239	2.5	2712	6 ABU39146	ABU39146 Protein e
42	238.5	2.5	2504	6 ADA34534	Ada34534 Acinetoba
43	238	2.5	2261	6 ABU18914	ABU18914 Pathogen
44	238	2.5	2271	6 ABM72734	ABM72734 Staphyloc
45	238	2.5	2283	6 ABP56876	ABP56876 Staphyloc

## ALIGNMENTS

### RESULT 1

AAW98011

ID AAW98011 standard; protein; 1838 AA.

XX AAW98011;

XX 21-JUN-1999 (first entry)

DE Hypersensitive response elicitor DspE.

XX Hypersensitive response elicitor; DspE; disease resistance;

KW insect resistance; biological control; transgenic plant.

XX Erwinia amylovora.

XX WO9907206-A1.

XX 18-FEB-1999.

XX 24-JUL-1998; 98WO-US015426.

XX 06-AUG-1997; 97US-0055105P.

XX (CORR ) CORNELL RES FOUND INC.

XX Bogdanove AJ, Kim JF, Wei Z, Beer SV;

XX WPI; 1999-180362/15.

XX N-PSDB; AAX24810.

PT Nucleic acid encoding hypersensitive response-eliciting protein - used to improve growth of plants and impart resistance to disease and insects.

XX Claim 18; Page 51-56; 75pp; English.

XX This polypeptide comprises the 198 kDa hypersensitive response elicitor protein DspE of Erwinia amylovora. The nucleotide sequence of the dsp region of E. amylovora strain Ea321 was determined using subclones of pCPP430. A two-gene operon was discovered comprising dspE (see AAX24810) and dspF (see AAX24811). The isolated dsp DNA molecules and encoded proteins can be used to impart disease resistance to plants, to enhance plant growth, and/or to control insects on plants. This is achieved by applying a hypersensitive response elicitor protein or polypeptide in a non-infectious form to plants or plant seeds, or by producing transgenic plants or plant seeds transformed with DNA encoding a hypersensitive response elicitor. Protection can be provided against a wide range of viruses, bacteria, fungi and insects, e.g. tobacco mosaic virus and tomato mosaic virus, Pseudomonas syringae, Xanthomonas campestris,

CC Fusarium oxysporum, Phytophthora infestans, armyworm, diamondback moth,  
 CC etc. The method avoids use of infectious agents or polluting chemicals.  
 CC Claimed transgenic plants are selected from alfalfa, rice, wheat, barley,  
 CC rye, cotton, sunflower, peanut, corn, potato, bean, pea, chichory,  
 CC lettuce, endive, cabbage, brussel sprout, sweet potato, beet, parsnip,  
 CC turnip, cauliflower, broccoli, turnip, radish, spinach, onion, garlic,  
 CC eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber,  
 CC apple, pear, melon, citrus, strawberry, grape, raspberry, pineapple,  
 CC soybean, tobacco, tomato, sorghum, sugarcane, Arabidopsis thaliana,  
 CC Sainpaulia, petunia, pelargonium, poinsettia, chrysanthemum, carnation  
 CC and zinnia  
 XX  
 SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 2; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSPQNAASLAAGKNGKMPRIHQ 60  
 DB 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSPQNAASLAAGKNGKMPRIHQ 60  
 QY 61 STAADGISAHQKKSFLRGCLGTTKFSRSAPQGGTTHSKGATLRDLARDGGETQH 120  
 DB 61 STAADGISAHQKKSFLRGCLGTTKFSRSAPQGGTTHSKGATLRDLARDGGETQH 120  
 QY 121 EAAAPDAALTRSGGVKRRNMDMAGRPVYKGGSGEDKVPYQKRLHQLNFGMRQTMLS 180  
 DB 121 EAAAPDAALTRSGGVKRRNMDMAGRPVYKGGSGEDKVPYQKRLHQLNFGMRQTMLS 180  
 QY 181 KWAHPASANAGDRLQSPHPIPSHHEIKKEPVGTSKATTAHADVEIAQEDDDSEFQQ 240  
 DB 181 KWAHPASANAGDRLQSPHPIPSHHEIKKEPVGTSKATTAHADVEIAQEDDDSEFQQ 240  
 QY 241 LHQORLARERENPPQPKLGVAIPISARFQPKLTAAVESVLEGTDTTQSPKLPQSMKGS 300  
 DB 241 LHQORLARERENPPQPKLGVAIPISARFQPKLTAAVESVLEGTDTTQSPKLPQSMKGS 300  
 QY 301 GAGVTPPLAVTLQKGLQALPDPNPPALNTLLKOTGLKDTQHYLAHASSDGSQHLLDNKG 360  
 DB 301 GAGVTPPLAVTLQKGLQALPDPNPPALNTLLKOTGLKDTQHYLAHASSDGSQHLLDNKG 360  
 QY 361 HLFDFIKSTATSVLHNSHPGEIKGLAQAGTGSVSDGSKGKISLGSTQSHNKTMLSQ 420  
 DB 361 HLFDFIKSTATSVLHNSHPGEIKGLAQAGTGSVSDGSKGKISLGSTQSHNKTMLSQ 420  
 QY 421 PGEAHRSLTGTWQHPAGAARQGESIRLHDDKIHLHPGLGWQSDAKDTHQSLSRAD 480  
 DB 421 PGEAHRSLTGTWQHPAGAARQGESIRLHDDKIHLHPGLGWQSDAKDTHQSLSRAD 480  
 QY 481 GKLYALKONRTQNLSDNKSSEKLVDKIKSVSDQGVAILTDTTPGRHKMSIMPSLDAS 540  
 DB 481 GKLYALKONRTQNLSDNKSSEKLVDKIKSVSDQGVAILTDTTPGRHKMSIMPSLDAS 540  
 QY 541 PESHISLSLHFAADAHGGLHSELEAQSVALSHGRLVVADESEKLFSAIIPKQDGNEL 600  
 DB 541 PESHISLSLHFAADAHGGLHSELEAQSVALSHGRLVVADESEKLFSAIIPKQDGNEL 600  
 QY 601 KXKAMPQALDEHFGHDHQISGFFHDDHGQNALVKNFRQQAACPLGNDHOFHPGWNLT 660  
 DB 601 KXKAMPQALDEHFGHDHQISGFFHDDHGQNALVKNFRQQAACPLGNDHOFHPGWNLT 660  
 QY 661 DALVIDNQGLHHTPEPHEIILDMHGLSLAQEGKLYHFDQLTGKWTGAESDCKQLKKG 720  
 DB 661 DALVIDNQGLHHTPEPHEIILDMHGLSLAQEGKLYHFDQLTGKWTGAESDCKQLKKG 720  
 QY 721 LDGAAYLLKQGVKRLNINQSTSSIKHGTEPVFSLPHVENKPEPGDALQGNKDKQAAM 780  
 DB 721 LDGAAYLLKQGVKRLNINQSTSSIKHGTEPVFSLPHVENKPEPGDALQGNKDKQAAM 780  
 QY 781 AVIGNVKYALTEKGDIRSFOIKPGTQQLERPAQTLREGISGELKDIHVHDKQNLAYLT 840  
 DB 781 AVIGNVKYALTEKGDIRSFOIKPGTQQLERPAQTLREGISGELKDIHVHDKQNLAYLT 840

RESULT 2  
 AAY71095  
 ID AAY71095 standard; protein; 1838 AA.

QY 841 HEGEVHQPREAWQNGAESSSWHKLALPQSESKLSLDMSEHKPIATPFDGSOHLKAG 900  
 DB 841 HEGEVHQPREAWQNGAESSSWHKLALPQSESKLSLDMSEHKPIATPFDGSOHLKAG 900  
 QY 901 GHAYAAPRGPIAVGTSGQTVFNRLMQGVKVIPIGSGLTVKLSAQTOGTMGAEGRKV 960  
 DB 901 GHAYAAPRGPIAVGTSGQTVFNRLMQGVKVIPIGSGLTVKLSAQTOGTMGAEGRKV 960  
 QY 961 SSKFSEIRIAYAFNPTMSTPRPIKNAAYATQHWQREGKPLIYEQCALIKQIDAHNVR 1020  
 DB 961 SSKFSEIRIAYAFNPTMSTPRPIKNAAYATQHWQREGKPLIYEQCALIKQIDAHNVR 1020  
 QY 1021 HNAQPDLQSKLETLDLGEGHGAELNDMKRFRDELEQSATRSVTVLGQHGVLKNGEIN 1080  
 DB 1021 HNAQPDLQSKLETLDLGEGHGAELNDMKRFRDELEQSATRSVTVLGQHGVLKNGEIN 1080  
 QY 1081 SEFKPSFGKALVQSFNNRSGQDLKSLOQAVHATPPSAESKLQSMGLGHFVSAGVDSHQ 1140  
 DB 1081 SEFKPSFGKALVQSFNNRSGQDLKSLOQAVHATPPSAESKLQSMGLGHFVSAGVDSHQ 1140  
 QY 1141 KGEIPLGRORDPNDKTALTKSRLILDTVTIGELHELADKAKIVSDHKPDADQIKQLRQOF 1200  
 DB 1141 KGEIPLGRORDPNDKTALTKSRLILDTVTIGELHELADKAKIVSDHKPDADQIKQLRQOF 1200  
 QY 1201 DTLREKRYESNPVCHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTTTTLVLESQ 1260  
 DB 1201 DTLREKRYESNPVCHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTTTTLVLESQ 1260  
 QY 1261 SAEALAKLQNTLLSLDSGESMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
 DB 1261 SAEALAKLQNTLLSLDSGESMFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
 QY 1321 LSFRTSGGLNVSGFDCGVGNIVATGHDVWYMTGKTSAGNASDMLSAKHKISPD 1380  
 DB 1321 LSFRTSGGLNVSGFDCGVGNIVATGHDVWYMTGKTSAGNASDMLSAKHKISPD 1380  
 QY 1381 RIGAAVSGTLQGTQNLKFKLTDELPFGIHLGTHGLTPAELLQKGIEHOMKQSKLT 1440  
 DB 1381 RIGAAVSGTLQGTQNLKFKLTDELPFGIHLGTHGLTPAELLQKGIEHOMKQSKLT 1440  
 QY 1441 FSVDTSANLDRAGINLNDGSKPVGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500  
 DB 1441 FSVDTSANLDRAGINLNDGSKPVGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500  
 QY 1501 ASNNRPTFLNGVAGANLTAALGVVAHSTHEGKPVGIPPAFTSTNVSAALADNRSTQSI 1560  
 DB 1501 ASNNRPTFLNGVAGANLTAALGVVAHSTHEGKPVGIPPAFTSTNVSAALADNRSTQSI 1560  
 QY 1561 SLELKRABPVTNSDISELTSTLKGKFKDSATTMKAALAKELDDAKPAEQHLIIQQHFSAK 1620  
 DB 1561 SLELKRABPVTNSDISELTSTLKGKFKDSATTMKAALAKELDDAKPAEQHLIIQQHFSAK 1620  
 QY 1621 DVVGDYERAEVNLKLVIRQQAADSHMELGASHSTTYNNLSRINNDGIVELLHKHFD 1680  
 DB 1621 DVVGDYERAEVNLKLVIRQQAADSHMELGASHSTTYNNLSRINNDGIVELLHKHFD 1680  
 QY 1681 AALPASSAKRLGEMNNDPALKDIIKQLQSTPSSASVSMELKDGLREOTEKAILDGKVG 1740  
 DB 1681 AALPASSAKRLGEMNNDPALKDIIKQLQSTPSSASVSMELKDGLREOTEKAILDGKVG 1740  
 QY 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSERNITGTFNFKYG 1800  
 DB 1741 REEVGVLFQDRNNLRVKSVSVSQSVSKSEGFNTPALLGTSNSAAMSERNITGTFNFKYG 1800  
 QY 1801 ODQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMKS 1838  
 DB 1801 ODQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMKS 1838

XX  
AC  
XX  
DT  
DE  
DE  
KW  
KW  
XX  
OS  
XX  
PN  
XX  
PD  
XX  
PF  
XX  
PR  
XX  
PA  
XX  
PI  
XX  
DR  
DR  
DR  
PT  
PT  
XX  
PS  
XX  
CC  
CC  
CC  
CC  
CC  
CC  
XX  
SQ

AAAY71095;  
08-SEP-2000 (first entry)  
Erwinia amylovora hypersensitive response elicitor encoded by dspE gene.  
Hypersensitive response elicitor; environmental stress resistance; plant;  
pathogen; dape gene.  
Erwinia amylovora.  
WO200028055-A2.  
18-MAY-2000.  
04-NOV-1999; 99WO-US026039.  
05-NOV-1998; 98US-0107243P.  
(EDEN-) EDEN BIOSCIENCE CORP.  
Wei Z, Schading RL;  
WPI; 2000-376566/32.  
N-PSDB; AAD00670.  
Application of a hypersensitive response elicitor protein to plants to  
impart stress resistance.  
Disclosure; Page 15-20; 84pp; English.  
The patent discloses a method to impart stress resistance to plants by  
applying a hypersensitive response elicitor in a non-infectious form to a  
plant or seed. The present sequence is a hypersensitive response elicitor  
encoded by dspE gene from Erwinia amylovora. The protein elicits plant  
pathogen's hypersensitive response and is used to impart stress  
resistance to plants  
Sequence 1838 AA;  
Query Match 100.0%; Score 9448; DB 3; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSSPQNAASLAASGKRGKMPRIHQ 60  
DB 1 MELKSLGTEHKAHVHTAAHNPVGHVALQOGSSSSSPQNAASLAASGKRGKMPRIHQ 60  
QY 61 STAADGISAHQOKKPSLRCGLTKKFSRQPOGPTTHSKGATLRDLIARDGETQH 120  
DB 61 STAADGISAHQOKKPSLRCGLTKKFSRQPOGPTTHSKGATLRDLIARDGETQH 120  
QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGSGEDKVPQOKRHOLNPNFGQRMQMLS 180  
DB 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGSGEDKVPQOKRHOLNPNFGQRMQMLS 180  
QY 181 KMAHPASAGDRLOHSPPHIPGSHHEIKKEPVGSTSKATTAHADRVIEAQEDDDSEFQ 240  
DB 181 KMAHPASAGDRLOHSPPHIPGSHHEIKKEPVGSTSKATTAHADRVIEAQEDDDSEFQ 240  
QY 241 LHQORLARERNPPQPKLGVA TPISARFQPKLTAVAESVLEGDTTOSPLKPOSMLKGS 300  
DB 241 LHQORLARERNPPQPKLGVA TPISARFQPKLTAVAESVLEGDTTOSPLKPOSMLKGS 300  
QY 301 GAGVTPLAVTLDKGKLQAPNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
DB 301 GAGVTPLAVTLDKGKLQAPNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
QY 361 HLFDIKSTATSVYLNHSHPGIEIKGLAQAGTGSVSDGSKGKISLGSGTQSHNKTMLSQ 420  
DB 361 HLFDIKSTATSVYLNHSHPGIEIKGLAQAGTGSVSDGSKGKISLGSGTQSHNKTMLSQ 420

QY 421 PGEAHRSLLTGWIQHHPAGARPQGESIRLHDDKIHLHPELGWQSQADKDTQSLSROAD 480  
DB 421 PGEAHRSLLTGWIQHHPAGARPQGESIRLHDDKIHLHPELGWQSQADKDTQSLSROAD 480  
QY 481 GKLYALKONRTLONLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTTPGRHMSIMPSLDAS 540  
DB 481 GKLYALKONRTLONLSDNKSSEKLVDKIKSYSVDQRGQVAILTDTTPGRHMSIMPSLDAS 540  
QY 541 PESHISLSLHFAHOGHLLHGKSELEAQAQVAISHGRVLVADSEGLFSAAPKQGDGNE 600  
DB 541 PESHISLSLHFAHOGHLLHGKSELEAQAQVAISHGRVLVADSEGLFSAAPKQGDGNE 600  
QY 601 KMKAMPQHALDEHFHGHQISGFPHDDHGLQNALVKNFRQHQHACPLGNDHQHFGHWNLT 660  
DB 601 KMKAMPQHALDEHFHGHQISGFPHDDHGLQNALVKNFRQHQHACPLGNDHQHFGHWNLT 660  
QY 661 DALVIDNOLGLHHTNPPEHILDMGHLSLALQEGKLIHYFDQLTGKWTGAESCKQLKKG 720  
DB 661 DALVIDNOLGLHHTNPPEHILDMGHLSLALQEGKLIHYFDQLTGKWTGAESCKQLKKG 720  
QY 721 LDGAAYLLKDGGEVKRLNINQSTSSIKHGTENVSFLPHVRNKPFGDALQGLNKKDAQAM 780  
DB 721 LDGAAYLLKDGGEVKRLNINQSTSSIKHGTENVSFLPHVRNKPFGDALQGLNKKDAQAM 780  
QY 781 AVIGVKNYLALTEKDIRSQIKPQGTQQLRPAQTLSREGISGELKDIDHVDHQNLYALT 840  
DB 781 AVIGVKNYLALTEKDIRSQIKPQGTQQLRPAQTLSREGISGELKDIDHVDHQNLYALT 840  
QY 841 HEGEVFHQPREAWONGAESSEWHKLALPQSEKLSLDMSHEHKPIATFEDGSHQOLKAG 900  
DB 841 HEGEVFHQPREAWONGAESSEWHKLALPQSEKLSLDMSHEHKPIATFEDGSHQOLKAG 900  
QY 901 GWIYAAPERGLAVGTSGSQTQVFNRLMOGVKGVIPGSGLTVKLSAQTCGTMGAERKV 960  
DB 901 GWIYAAPERGLAVGTSGSQTQVFNRLMOGVKGVIPGSGLTVKLSAQTCGTMGAERKV 960  
QY 961 SSKFSERIRAYANPNWSTPRPIKNAAYATQHGWQREGKLPKLYEMOGALIKQJDAHNVR 1020  
DB 961 SSKFSERIRAYANPNWSTPRPIKNAAYATQHGWQREGKLPKLYEMOGALIKQJDAHNVR 1020  
QY 1021 HNAPODLOSKLETLDLGEHGAELLNDMKRPRELEQSATRSVTVLGQHGVLSKNGEIN 1080  
DB 1021 HNAPODLOSKLETLDLGEHGAELLNDMKRPRELEQSATRSVTVLGQHGVLSKNGEIN 1080  
QY 1081 SEFKPSGKALVQSFVNVNRSQDLSSKLSQAQVHATPFAESKLSQMLGHFVSAGVDMHQ 1140  
DB 1081 SEFKPSGKALVQSFVNVNRSQDLSSKLSQAQVHATPFAESKLSQMLGHFVSAGVDMHQ 1140  
QY 1141 KGEIPLGRQDPNDKTALTKSRLILDVTITGELHELADKAKLVSDHKPDADQIKQLRQOF 1200  
DB 1141 KGEIPLGRQDPNDKTALTKSRLILDVTITGELHELADKAKLVSDHKPDADQIKQLRQOF 1200  
QY 1201 DTUREKYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKXHHGVNLTTRTVLESQ 1260  
DB 1201 DTUREKYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKXHHGVNLTTRTVLESQ 1260  
QY 1261 SAEIAXKLLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
DB 1261 SAEIAXKLLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
QY 1321 LSFRTSGGLNVSPGRGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHISPD 1380  
DB 1321 LSFRTSGGLNVSPGRGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHISPD 1380  
QY 1381 RIGAAVSGTLOCTLONSLKFKLDEDELPFGTHGTLTPAELLKXGIEHOMQGSKLT 1440  
DB 1381 RIGAAVSGTLOCTLONSLKFKLDEDELPFGTHGTLTPAELLKXGIEHOMQGSKLT 1440  
QY 1441 FSVDTSANLDRAGINLNDGSKPVGTVARVASGLSASANLAAGSRERSTTSGFGSTTS 1500  
DB 1441 FSVDTSANLDRAGINLNDGSKPVGTVARVASGLSASANLAAGSRERSTTSGFGSTTS 1500  
QY 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAPTSTNVSAALADNRTSOSI 1560

1501 ASNNRPTFLNGVAGANLTAALGVHSSTHEGKPVGIFPAFTSTNSAALDNRTSQSI 1560  
 1561 SLELKRAEPTVNDISBLTSTLGHFKDSTATTWLAALKELDDAKPAEQHLHILOQHPSAK 1620  
 1561 SLELKRAEPTVNDISBLTSTLGHFKDSTATTWLAALKELDDAKPAEQHLHILOQHPSAK 1620  
 1621 DVVGDEREYAVRNLLKLVIRQQAADSHSMELGSAHSHTYNNLSRINNDGIVELLKHFD 1680  
 1621 DVVGDEREYAVRNLLKLVIRQQAADSHSMELGSAHSHTYNNLSRINNDGIVELLKHFD 1680  
 1681 AALPASSAKELGEMMNDPALKDIIKLOLSTPSSASVSVMELKDGLEQTEKAILDGKVG 1740  
 1681 AALPASSAKELGEMMNDPALKDIIKLOLSTPSSASVSVMELKDGLEQTEKAILDGKVG 1740  
 1741 REEVGVLFQDRNNLRVKSVSVOSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800  
 1741 REEVGVLFQDRNNLRVKSVSVOSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800  
 1801 QDQNTPRFTLEGGIAQANPQVASALTDLKKEGLEWKS 1838  
 1801 QDQNTPRFTLEGGIAQANPQVASALTDLKKEGLEWKS 1838

RESULT 3

AA184856

ID AA184856 standard; protein; 1838 AA.

AC AC

XX AA184856;

XX 08-AUG-2000 (first entry)

XX A hypersensitive response elicitor protein.

XX Hypersensitive response; insect control; disease resistance;

XX KW hypersensitive response elicitor; plant growth; vegetable; crop;

XX KW ornamental plant; dspE gene.

XX OS Erwinia amylovora.

XX PN WO200020452-A2.

XX 13-APR-2000.

XX PF 05-OCT-1999; 99WO-US023181.

XX PR 05-OCT-1998; 98US-0103050P.

XX PA (EDEN-) EDEN BIOSCIENCE CORP.

XX PI Wei Z, Fan H, Niggemeyer JL;

XX DR WPI; 2000-303745/26.

XX DR N-PSDB; AAA14940.

XX Hypersensitive response elicitor polypeptides useful for imparting

XX PT enhanced growth, disease resistance and insect resistance to plants,

XX PT especially vegetables and ornamental flowers.

XX PS Disclosure; Page 17-22; 100pp; English.

XX The present sequence represents a hypersensitive response elicitor

XX CC polypeptide. the polynucleotide represents the dspE gene. The

XX CC specification describes hypersensitive response elicitor polypeptide

XX CC fragments, which do not elicit a hypersensitive response. Instead, the

XX CC proteins impart disease resistance to plants, enhance plant growth,

XX CC and/or control insects. The polypeptide fragments may be used to these

XX CC properties to plants. The plants which may be treated in this way include

XX CC vegetables, crops and ornamental plants such as alfalfa, rice, wheat,

XX CC barley, rye, cotton, sunflower, peanut, corn, potato, sweet potato, bean,

XX CC pea, chichory, lettuce, endive, cabbage, brussel sprout, beet, parsnip,

XX CC turnip, cauliflower, broccoli, radish, spinach, onion, garlic, eggplant,

XX CC pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear,

CC melon, citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,  
 CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia, petunia,  
 CC pelargonium, poinsettia, chrysanthemum, carnation or zinnia  
 XX  
 SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 3; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVHTAAHNPVGHVALQGGSSSSPQNAASIAAEGKRGKMPRIHQ 60  
 Db 1 MELKSLGTEHKAHVHTAAHNPVGHVALQGGSSSSPQNAASIAAEGKRGKMPRIHQ 60  
 QY 61 STAADGISAHQKQKFSLRGCLGTTKFSRSPQGGPQGGTTHSGATLRDLARDDDGETQH 120  
 Db 61 STAADGISAHQKQKFSLRGCLGTTKFSRSPQGGPQGGTTHSGATLRDLARDDDGETQH 120  
 QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVPTQKRLHNNFGQMTLS 180  
 Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGGSGEDKVPTQKRLHNNFGQMTLS 180  
 QY 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGVSTSKATTAHADRVETAQEDDDSEFQ 240  
 Db 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPPVGVSTSKATTAHADRVETAQEDDDSEFQ 240  
 QY 241 LHOQRLARERENPPQPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKXQSMKGS 300  
 Db 241 LHOQRLARERENPPQPKLGVA TPISARFQPKLTAVAESVLEGTDTTQSPKXQSMKGS 300  
 QY 301 GAGVTP LAVTLDKGLQ LAPDNP PALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
 Db 301 GAGVTP LAVTLDKGLQ LAPDNP PALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
 QY 361 HLFDIKSTATSYVLHNSHPGRIKGLAQAGTGSVSDGSKGILSGSTQSHNKTLSQ 420  
 Db 361 HLFDIKSTATSYVLHNSHPGRIKGLAQAGTGSVSDGSKGILSGSTQSHNKTLSQ 420  
 QY 421 PGEAHRSLTIGIWHQHPAGAAARPGESIRLHDDKIHLHPGLVWQSAADKDTHSLSROAD 480  
 Db 421 PGEAHRSLTIGIWHQHPAGAAARPGESIRLHDDKIHLHPGLVWQSAADKDTHSLSROAD 480  
 QY 481 GKLYALKONRTLQNLSDNKSSEKLVDTKSYSDQGVAILTDTTPGRHKMSIMPSLDAS 540  
 Db 481 GKLYALKONRTLQNLSDNKSSEKLVDTKSYSDQGVAILTDTTPGRHKMSIMPSLDAS 540  
 QY 541 PESHISLSLHFAADAHQGLLHGKSELEAQSVAISHGRLVVADSEGLKLSAAIPKQGGNEL 600  
 Db 541 PESHISLSLHFAADAHQGLLHGKSELEAQSVAISHGRLVVADSEGLKLSAAIPKQGGNEL 600  
 QY 601 KMKAMPQHALDRHFHGHQHSIGFFHDDHQLNALVKNFRQOQHACPLGNDHQHPGWNLT 660  
 Db 601 KMKAMPQHALDRHFHGHQHSIGFFHDDHQLNALVKNFRQOQHACPLGNDHQHPGWNLT 660  
 QY 661 DALVTDNQLGLHHTNPEPHEILDMGHLSLALQEGKLYHFDQLTGKWTGAESDCKQLKKG 720  
 Db 661 DALVTDNQLGLHHTNPEPHEILDMGHLSLALQEGKLYHFDQLTGKWTGAESDCKQLKKG 720  
 QY 721 LDGAAYLLKDGKVRKLNINQSTSSIKHGTEVPSLPHVRNKPDPGALQGLNKKDAQAM 780  
 Db 721 LDGAAYLLKDGKVRKLNINQSTSSIKHGTEVPSLPHVRNKPDPGALQGLNKKDAQAM 780  
 QY 781 AVIGVKNYALATEKGDIRSFOIKPQTQQLERPAQTLREGISGELKD IHVDHKQNLALT 840  
 Db 781 AVIGVKNYALATEKGDIRSFOIKPQTQQLERPAQTLREGISGELKD IHVDHKQNLALT 840  
 QY 841 HEGEVFHQPREAWQNGAESSSWHKIALPQSESKLSLDMSHHEHKPIATPDSGSHQOLKAG 900  
 Db 841 HEGEVFHQPREAWQNGAESSSWHKIALPQSESKLSLDMSHHEHKPIATPDSGSHQOLKAG 900  
 QY 901 GWHVAAAPERGPLAVGTSGSTVFNRMLQGVKGVKIPGSLTVKLSAQTGGMTGAEGRKV 960  
 Db 901 GWHVAAAPERGPLAVGTSGSTVFNRMLQGVKGVKIPGSLTVKLSAQTGGMTGAEGRKV 960



QY 961 SSKFSRIRAYAFNPMTSPRIKNAAYATQHGMQREGKPLVEMOGALIKOLDAHNVR 1020  
Db 961 SSKFSRIRAYAFNPMTSPRIKNAAYATQHGMQREGKPLVEMOGALIKOLDAHNVR 1020  
QY 1021 HNAPODLSQKLETLDELGEHGAELNDMKRFRDELEQSATRSVTVLGQHGVLSNGEIN 1080  
Db 1021 HNAPODLSQKLETLDELGEHGAELNDMKRFRDELEQSATRSVTVLGQHGVLSNGEIN 1080  
QY 1081 SEFKPSGKALVOSFNVRSGDLSKLSLOAIVATPPSAESKLSMLGHFVSAGVMSHQ 1140  
Db 1081 SEFKPSGKALVOSFNVRSGDLSKLSLOAIVATPPSAESKLSMLGHFVSAGVMSHQ 1140  
QY 1141 KGRIPIGRQDNDKTKALTSRLILDTVTIGELHELADKAKLYSDHKPDADQIKQLRQOF 1200  
Db 1141 KGRIPIGRQDNDKTKALTSRLILDTVTIGELHELADKAKLYSDHKPDADQIKQLRQOF 1200  
QY 1201 DTLREKRYESNPVKHYTDMGFTINKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260  
Db 1201 DTLREKRYESNPVKHYTDMGFTINKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260  
QY 1261 SAEKAKLNTLISLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVPIGAGITLDRAIN 1320  
Db 1261 SAEKAKLNTLISLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVPIGAGITLDRAIN 1320  
QY 1321 LSPSRTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380  
Db 1321 LSPSRTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380  
QY 1381 RIGAAVSGTLQGTLSNPKLTELDPGFIHGLTHGTLTPAELLQKGIHQMKGSKLT 1440  
Db 1381 RIGAAVSGTLQGTLSNPKLTELDPGFIHGLTHGTLTPAELLQKGIHQMKGSKLT 1440  
QY 1441 FSVDTSANLDRAGINLNDGSPGVTVARVSAGLSASANLAAGSRRTSTTSGQFGSTTS 1500  
Db 1441 FSVDTSANLDRAGINLNDGSPGVTVARVSAGLSASANLAAGSRRTSTTSGQFGSTTS 1500  
QY 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
Db 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
QY 1561 SLEKRAEPTVNSDISELTSLGKHFKDSATTKMLAALKELDDAKPAEQLHILQHFSAK 1620  
Db 1561 SLEKRAEPTVNSDISELTSLGKHFKDSATTKMLAALKELDDAKPAEQLHILQHFSAK 1620  
QY 1621 DVVGDREYEAURNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680  
Db 1621 DVVGDREYEAURNLKKLVIRQQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHKHFD 1680  
QY 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPSSASVSMELKDLREQTEKAILDGKVG 1740  
Db 1681 AALPASSAKRLGEMMNNDPALKDIIKQLQSTPSSASVSMELKDLREQTEKAILDGKVG 1740  
QY 1741 REEVGVLFDQRNNLRVKSVSQSVKSEGFNTPALLLGTSNAAEMERNIGTINPKYG 1800  
Db 1741 REEVGVLFDQRNNLRVKSVSQSVKSEGFNTPALLLGTSNAAEMERNIGTINPKYG 1800  
QY 1801 QDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMS 1838  
Db 1801 QDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMS 1838

RESULT 4  
AAU02878  
ID AAU02878 standard; protein; 1838 AA.  
XX  
AC AAU02878;  
XX  
DT 23-OCT-2001 (first entry)  
XX  
DE Erwinia amylovora disease-specific region E (dspE) polypeptide.  
XX  
KW Disease-specific region E; dspE; insect control; disease resistance;

hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic; cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr; petunia; chrysanthemum; carnation; transgenic plant; avirulence locus; pathogenicity; fireblight.  
XX Erwinia amylovora.  
XX OS  
XX US6228644-B1.  
XX  
XX PD 08-MAY-2001.  
XX PF 22-JUL-1998; 98US-00120663.  
XX PR 06-AUG-1997; 97US-0055106P.  
XX PA (CORR ) CORNELL RES FOUND INC.  
XX PI Bogdanove AJ, Kim JF, Wei Z, Beer SV;  
XX DR WPI; 2001-327491/34.  
XX DR N-PSDB; AAS05199.  
XX  
PT Polynucleotides encoding hypersensitive response eliciting proteins or polypeptides useful for imparting disease resistance to plants, to enhance plant growth, and/or to control insects on plants.  
PS Claim 1; Col 9-18; 37pp; English.  
XX  
CC The sequence represents an Erwinia amylovora disease-specific region E (dspE) polypeptide, which elicits a hypersensitive response in plants. The dspE operon functions as an avirulence (avr) locus and the dspE protein is required for pathogenicity of Erwinia amylovora. DspE is also responsible for causing fireblight in certain plants. The protein and its associated nucleic acid can be applied in a non-infectious form to plants or plant seeds to impart disease resistance, to enhance plant growth, and/or to control insects on plants. Alternatively, plant cells may be transformed with the DNA to form transgenic plants with the same properties. The method can be utilised to treat a wide variety of plants and seeds, including crop plants such as rice, wheat, corn, cabbage, cauliflower, garlic, cucumber, apple, grape, tobacco and sugarcane, and ornamental plants such as Arabidopsis thaliana, petunia, chrysanthemum and carnation  
XX  
SQ Sequence 1838 AA;  
Query Match 100.0%; Score 9448; DB 4; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MELKSLGTEHKAHVHTAAHNPVGHVVALQGGSSSSPQNAASLAARGNKRGKMPRIHQ 60  
Db 1 MELKSLGTEHKAHVHTAAHNPVGHVVALQGGSSSSPQNAASLAARGNKRGKMPRIHQ 60  
QY 61 STAADGISAAHQKQKFSLRGCLGTTKFSRSAPQGGPQTHSGATLRLDARDGTOH 120  
Db 61 STAADGISAAHQKQKFSLRGCLGTTKFSRSAPQGGPQTHSGATLRLDARDGTOH 120  
QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGSGEDKVPQOKRHLNNGFQOMRTMLS 180  
Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGSGEDKVPQOKRHLNNGFQOMRTMLS 180  
QY 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPVSGTSKATTAHADRVETIAQEDDDSEFQ 240  
Db 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKEEPVSGTSKATTAHADRVETIAQEDDDSEFQ 240  
QY 241 LHOORLARENPPOPKLGVATPI SARFOPKLTAVAESVLEGTDTTQSPKPKQSMKGS 300  
Db 241 LHOORLARENPPOPKLGVATPI SARFOPKLTAVAESVLEGTDTTQSPKPKQSMKGS 300  
QY 301 GAGVTPLAVTLDKGKQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360  
Db 301 GAGVTPLAVTLDKGKQLAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKG 360

QY 361 HLFIDIKSTATSVLHNSHPGIBKGLAQAGTGSVVDGSGKISLGSTQGHNTMLSQ 420  
 Db 361 HLFIDIKSTATSVLHNSHPGIBKGLAQAGTGSVVDGSGKISLGSTQGHNTMLSQ 420  
 QY 421 PGEAHRSLTGTWHPAGAAPQGESIRUHDHDKIHILHPELGVWQSDKTHQSQRQAD 480  
 Db 421 PGEAHRSLTGTWHPAGAAPQGESIRUHDHDKIHILHPELGVWQSDKTHQSQRQAD 480  
 QY 481 GKLYALKDNRTLQNTSDNKSSEKLVDKIKSYSDVQGOVAIITDTPGRHKMSIMPSLDAS 540  
 Db 481 GKLYALKDNRTLQNTSDNKSSEKLVDKIKSYSDVQGOVAIITDTPGRHKMSIMPSLDAS 540  
 QY 541 PESHISLSLHFAHAGHGLHGKSELEAQSAVISHGRLVWADSEGKLFSAAIKQGDGDEL 600  
 Db 541 PESHISLSLHFAHAGHGLHGKSELEAQSAVISHGRLVWADSEGKLFSAAIKQGDGDEL 600  
 QY 601 KMKAMPQHALLDHFGHGHQISGFFHDDHGLNALVKNFRQOAHACPLGNDHGFHPGWNLT 660  
 Db 601 KMKAMPQHALLDHFGHGHQISGFFHDDHGLNALVKNFRQOAHACPLGNDHGFHPGWNLT 660  
 QY 661 DALVIDNQLGLHHTPEPEHEIILDMHGLSLAQEGKLYHFDQITKGWTAESDCQKQKKG 720  
 Db 661 DALVIDNQLGLHHTPEPEHEIILDMHGLSLAQEGKLYHFDQITKGWTAESDCQKQKKG 720  
 QY 721 LDGAAYLLKDXGEVKRLININQSTSSIKHGTENVPSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 Db 721 LDGAAYLLKDXGEVKRLININQSTSSIKHGTENVPSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 QY 781 AVIGUNKYIALTEKDIRSQIKPGTQQLERPAQTLRSREGISGELKDIIHVDHKQNIYALT 840  
 Db 781 AVIGUNKYIALTEKDIRSQIKPGTQQLERPAQTLRSREGISGELKDIIHVDHKQNIYALT 840  
 QY 841 HGEVFPHPREAWQNGAESSEHKLALPOSESKLSLDMSHHEKPTATFEDSGSQHOLKAG 900  
 Db 841 HGEVFPHPREAWQNGAESSEHKLALPOSESKLSLDMSHHEKPTATFEDSGSQHOLKAG 900  
 QY 901 GWEHAYAAPRGPLAVGTSGSTVFNRMLMQGVKVIPIGSLTVKLSAQTCGMGTGAEGRKV 960  
 Db 901 GWEHAYAAPRGPLAVGTSGSTVFNRMLMQGVKVIPIGSLTVKLSAQTCGMGTGAEGRKV 960  
 QY 961 SKKFSERIRAYAPNPTMSTPRPKNAAYATQGWQREGKLPLEYMOGALLIKQLDAHNV 1020  
 Db 961 SKKFSERIRAYAPNPTMSTPRPKNAAYATQGWQREGKLPLEYMOGALLIKQLDAHNV 1020  
 QY 1021 HNAPODLSQKLETLDLGHEGAEELNDMKRFRDELEQSATRSVTVLQGHQVGLKNGEIN 1080  
 Db 1021 HNAPODLSQKLETLDLGHEGAEELNDMKRFRDELEQSATRSVTVLQGHQVGLKNGEIN 1080  
 QY 1081 SEFKPSPGKALVQSFNVNRSQDLSKSLQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 Db 1081 SEFKPSPGKALVQSFNVNRSQDLSKSLQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 QY 1141 KGEIPLGRORDNDKALTAKSLRLITDVTIGELHELADKAKLVSDHKPDADQIKQLROOF 1200  
 Db 1141 KGEIPLGRORDNDKALTAKSLRLITDVTIGELHELADKAKLVSDHKPDADQIKQLROOF 1200  
 QY 1201 DTLREKRYESNPVXHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260  
 Db 1201 DTLREKRYESNPVXHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQG 1260  
 QY 1261 SHELAKLNTLLSDSGESMFSRSYGGGVSTVFVPTLSKVPVPVPIPCAGITLDRAVN 1320  
 Db 1261 SHELAKLNTLLSDSGESMFSRSYGGGVSTVFVPTLSKVPVPVPIPCAGITLDRAVN 1320  
 QY 1321 LSFSTSGGLNVSGFRDGGVGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380  
 Db 1321 LSFSTSGGLNVSGFRDGGVGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPD 1380  
 QY 1381 RIGAAVSGTLQGTQNSLKFKTEDELPGFIIHGLTHGTLTPAELLQKGIHEQMKGSKLT 1440  
 Db 1381 RIGAAVSGTLQGTQNSLKFKTEDELPGFIIHGLTHGTLTPAELLQKGIHEQMKGSKLT 1440  
 QY 1441 FSVDTISANLDRAGINLNEGSKPNGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500

Db 1441 FSVDTISANLDRAGINLNEGSKPNGVTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500  
 QY 1501 ASNNRPTFLNGVGAGANLTAALGVVAHSSTHEGKPVGIFPAFTSTNVSAALADNRTSQSI 1560  
 Db 1501 ASNNRPTFLNGVGAGANLTAALGVVAHSSTHEGKPVGIFPAFTSTNVSAALADNRTSQSI 1560  
 QY 1561 SLELKRAEPTVNDISELTSTLTKGHPKDSATTKMLAALKELDDAKPAEQHLILQOHFSAK 1620  
 Db 1561 SLELKRAEPTVNDISELTSTLTKGHPKDSATTKMLAALKELDDAKPAEQHLILQOHFSAK 1620  
 QY 1621 DVVGDERYEAVRNKLKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLKHFD 1680  
 Db 1621 DVVGDERYEAVRNKLKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLKHFD 1680  
 QY 1681 AALPASSAKRLGEMMNDPALKDIIKQLOSTFPSSASVSMELKDGIREQTEKAILDGKVG 1740  
 Db 1681 AALPASSAKRLGEMMNDPALKDIIKQLOSTFPSSASVSMELKDGIREQTEKAILDGKVG 1740  
 QY 1741 REEVGVLFQDRNLRVKSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800  
 Db 1741 REEVGVLFQDRNLRVKSVSQSVSKSEGFNTPALLLGTSNSAAMSMERNIGTINFKYG 1800  
 QY 1801 QDQNTPRFTLEGGAQAQNPQVASALTDLKKEGLEWKS 1838  
 Db 1801 QDQNTPRFTLEGGAQAQNPQVASALTDLKKEGLEWKS 1838  
 RESULT 5  
 AAE16449  
 ID AAE16449 standard; protein; 1838 AA.  
 XX  
 AC AAE16449;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Erwinia amylovora hypersensitive response elicitor protein, dspE.  
 XX  
 KW Hypersensitive response elicitor protein; plant growth; fruit coloration;  
 KW disease resistance; stress resistance; phytotoxin; insect infection;  
 KW plant maturation; dspE protein.  
 XX  
 OS Erwinia amylovora.  
 XX  
 PN WO200198501-A2.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 12-JUN-2001; 2001WO-US018820.  
 XX  
 PR 16-JUN-2000; 2000US-0212211P.  
 XX  
 PA (EDEN-) EDEN BIOSCIENCE CORP.  
 XX  
 PI Fan H, Wei Z;  
 XX  
 DR WPI; 2002-122282/16.  
 DR N-PSDB; AAD27017.  
 XX  
 PT New hypersensitive response elicitor proteins comprising spaced apart  
 PT domains having an acidic portion linked to an alpha-helix, useful for  
 PT imparting disease or stress resistance, controlling insects or enhancing  
 PT plant growth.  
 XX  
 PS Disclosure; Page 18-23; 99pp; English.  
 XX  
 CC The patent discloses hypersensitive response elicitor proteins and  
 CC nucleotides encoding such proteins. Hypersensitive response elicitor  
 CC proteins comprise an isolated pair or more of spaced apart domains, each  
 CC comprising an acidic portion linked to an alpha-helix and capable of  
 CC eliciting a hypersensitive response in plants. Sequences of the invention  
 CC are used to impart disease resistance to plants, to enhance plant growth,  
 CC to control insects and/or to impart stress resistance to plants which

CC includes resistance to environmental stresses such as climate, air  
 CC pollution, chemical and nutritional stress. The method of imparting  
 CC disease resistance has the potential for treating previously untreatable  
 CC diseases, treating diseases systemically and avoiding the use of  
 CC infectious agents or environmentally harmful materials. Hyper- sensitive  
 CC response elicitor sequences are used to enhance plant growth which  
 CC encompasses greater yield, increased in quantity of seeds produced,  
 CC percentage of seeds germinated, plant size and biomass, bigger fruits,  
 CC earlier fruit coloration and plant maturation. They are also used for  
 CC insect control which encompasses preventing direct insect damage to plant  
 CC by feeding injury, interfering with insect larval feeding on the plants,  
 CC preventing insects from colonising host plants and releasing phytotoxins.  
 CC Sequences of the invention also prevent subsequent disease damage to  
 CC plants resulting from insect infection. The present sequence is Erwinia  
 CC amylovora hypersensitive response elicitor protein, dspe  
 XX  
 SQ Sequence 1838 AA;

Query Match 100.0%; Score 9448; DB 5; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELKSLGTEHKAHVTAHNPVGHVVALQQGSSSSSPQNAASLAAGKRGKMPRIHOP 60  
 Db 1 MELKSLGTEHKAHVTAHNPVGHVVALQQGSSSSSPQNAASLAAGKRGKMPRIHOP 60

Qy 61 STAADGISAHQKSPSLRGCLGCTKFSRPAQOGPTGTHSKGATLRDLARDGETQH 120  
 Db 61 STAADGISAHQKSPSLRGCLGCTKFSRPAQOGPTGTHSKGATLRDLARDGETQH 120

Qy 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVKGSGEDKVPQQKRRHQLNFGQMRQTMLS 180  
 Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVKGSGEDKVPQQKRRHQLNFGQMRQTMLS 180

Qy 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKPEPVGSGTSKATTAHADRVIEIAQEDDDSEFQ 240  
 Db 181 KMAHPASANAGDRLOHSPPHIPGSHHEIKPEPVGSGTSKATTAHADRVIEIAQEDDDSEFQ 240

Qy 241 LHQRLARERNPPQPKLGVAATPISARFQPKLTAVAESVLEGGTDTQSPKPKPSMLKGS 300  
 Db 241 LHQRLARERNPPQPKLGVAATPISARFQPKLTAVAESVLEGGTDTQSPKPKPSMLKGS 300

Qy 301 GAGVTPPLAVTLDKGKLOLAPNPPALNTLLKQTLQKDTQHYLAHSSDGSQHLLDNKG 360  
 Db 301 GAGVTPPLAVTLDKGKLOLAPNPPALNTLLKQTLQKDTQHYLAHSSDGSQHLLDNKG 360

Qy 361 HLFDIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTQSHNKTMLSQ 420  
 Db 361 HLFDIKSTATSYSLHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTQSHNKTMLSQ 420

Qy 421 PGEAHSLLTGIWOHPAGARPQGESIRLHDDKTHILHPELVGWSADKQTHSOLSRQAD 480  
 Db 421 PGEAHSLLTGIWOHPAGARPQGESIRLHDDKTHILHPELVGWSADKQTHSOLSRQAD 480

Qy 481 GKLYALKDNRTLQNLSDNKSSEKLVDKISKYSVDQRGQVAILTDPGRHKMSIMPSLDAS 540  
 Db 481 GKLYALKDNRTLQNLSDNKSSEKLVDKISKYSVDQRGQVAILTDPGRHKMSIMPSLDAS 540

Qy 541 PESHISLSLHFADAHQGLLHGKSLBAQSVASISGRVLVADSEGLKLSAAIPKQGDGNEI 600  
 Db 541 PESHISLSLHFADAHQGLLHGKSLBAQSVASISGRVLVADSEGLKLSAAIPKQGDGNEI 600

Qy 601 KMKAMPQHALDEHFGHDIQISGFFHDDHGQLNALVKNNFRQHQACPLGNDHQFHPGNLT 660  
 Db 601 KMKAMPQHALDEHFGHDIQISGFFHDDHGQLNALVKNNFRQHQACPLGNDHQFHPGNLT 660

Qy 661 DALVIDNQGLHHTNPPEHETLDMGHLGSLAQEGKLHYFDQLTKGWTGAESDCKQLKKG 720  
 Db 661 DALVIDNQGLHHTNPPEHETLDMGHLGSLAQEGKLHYFDQLTKGWTGAESDCKQLKKG 720

Qy 721 LDGAAYLLKQGEVRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 Db 721 LDGAAYLLKQGEVRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780

Qy 781 AVIGWNKYALTEKDIRSFOIKPGTQOLRPAOTLSREGISGELKDIHVDHKQNLVALT 840  
 Db 781 AVIGWNKYALTEKDIRSFOIKPGTQOLRPAOTLSREGISGELKDIHVDHKQNLVALT 840

Qy 841 HEGEVFHQPREAMONGAESSWHKLALPQSESKLSIDMSHEHKPIATTFEDGSHQLKAG 900  
 Db 841 HEGEVFHQPREAMONGAESSWHKLALPQSESKLSIDMSHEHKPIATTFEDGSHQLKAG 900

Qy 901 GWHAYAAPRGPLAVGTSGSQTENRMLMOGVKGVIPSGGLTVKLSAQTQGMTGAERGK 960  
 Db 901 GWHAYAAPRGPLAVGTSGSQTENRMLMOGVKGVIPSGGLTVKLSAQTQGMTGAERGK 960

Qy 961 SSKFESERIRAYAFNPTMTSTPRPIKNAAYATQHGQWQREGKPLKPLYEMOGALIKQIDAHNVR 1020  
 Db 961 SSKFESERIRAYAFNPTMTSTPRPIKNAAYATQHGQWQREGKPLKPLYEMOGALIKQIDAHNVR 1020

Qy 1021 HNAPOPLOKLETLDLGEHGAELNNDKMPRDELEQSATRSVTVLGQHGQVLKNSGEIN 1080  
 Db 1021 HNAPOPLOKLETLDLGEHGAELNNDKMPRDELEQSATRSVTVLGQHGQVLKNSGEIN 1080

Qy 1081 SEFKPSGKALVQSFNVRSGODLSKSLQQAQVHATPPSAESKLSQSMGLGHFVSAGVDMSHQ 1140  
 Db 1081 SEFKPSGKALVQSFNVRSGODLSKSLQQAQVHATPPSAESKLSQSMGLGHFVSAGVDMSHQ 1140

Qy 1141 KGEIPLGRQDPNDKTALTCSRILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1200  
 Db 1141 KGEIPLGRQDPNDKTALTCSRILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1200

Qy 1201 DTLEKKEYSNPVHYTDMGFTNKALEANYDAVKAFINAPKKEHHGVNLTTRTVLESQ 1260  
 Db 1201 DTLEKKEYSNPVHYTDMGFTNKALEANYDAVKAFINAPKKEHHGVNLTTRTVLESQ 1260

Qy 1261 SAEIPLGRQDPNDKTALTCSRILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1320  
 Db 1261 SAEIPLGRQDPNDKTALTCSRILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1320

Qy 1321 LSFRTSGGLNVSFGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHISPD 1380  
 Db 1321 LSFRTSGGLNVSFGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHISPD 1380

Qy 1381 RIGAAVSGTLQGTLONSLKFTEDELPGFTHGLTHGTLTPAEILLQKQIEHOMKQSGKLT 1440  
 Db 1381 RIGAAVSGTLQGTLONSLKFTEDELPGFTHGLTHGTLTPAEILLQKQIEHOMKQSGKLT 1440

Qy 1441 FSDVTSANLDIRAGINLEDGSKPVGTVARVAGLSASANLAAGSRSTSGFGSTTS 1500  
 Db 1441 FSDVTSANLDIRAGINLEDGSKPVGTVARVAGLSASANLAAGSRSTSGFGSTTS 1500

Qy 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSOSI 1560  
 Db 1501 ASNNRPTFLNGVGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALADNRTSOSI 1560

Qy 1561 SLELKRAEPTVNSDISETLSTLGHFKDQSAATKMLAALKELDLDDAKPAEQLHILQHFSAK 1620  
 Db 1561 SLELKRAEPTVNSDISETLSTLGHFKDQSAATKMLAALKELDLDDAKPAEQLHILQHFSAK 1620

Qy 1621 DVVGDREYAVRNLLKLVIRQQAADSHMELGSAHSTTVNNLSRINNDGIVELLHKHFD 1680  
 Db 1621 DVVGDREYAVRNLLKLVIRQQAADSHMELGSAHSTTVNNLSRINNDGIVELLHKHFD 1680

Qy 1681 AALPASSAKRLGEMNNNDPALKDIKOLQSTPSSASVSMELKDLREQTEKALDGVK 1740  
 Db 1681 AALPASSAKRLGEMNNNDPALKDIKOLQSTPSSASVSMELKDLREQTEKALDGVK 1740

Qy 1741 REEVGVLFQDRNNLRVKSVSVSGSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800  
 Db 1741 REEVGVLFQDRNNLRVKSVSVSGSEGFNTPALLGTSNSAAMSMERNIGTINFKYG 1800

Qy 1801 QDQNTPRRFTLEGGIAQANFQVSAALTDLKEGLEMS 1838  
 Db 1801 QDQNTPRRFTLEGGIAQANFQVSAALTDLKEGLEMS 1838

```

RESULT 6
AAU04490
ID AAU04490 standard; protein; 518 AA.
XX
AC AAU04490;
XX
DT 23-OCT-2001 (first entry)
XX
DE Erwinia amylovora disease-specific region E (dspE) mutant protein #1.
XX
KW Disease-specific region E; dspE; insect control; disease resistance;
KW hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic;
KW cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr;
KW petunia; chrysanthemum; carnation; transgenic plant; avirulence locus;
KW pathogenicity; fireblight; mutant; mutein.
XX
OS Erwinia amylovora.
OS Synthetic.
XX
PN US6228644-B1.
XX
PD 08-MAY-2001.
XX
PF 22-JUL-1998; 98US-00120663.
XX
PR 06-AUG-1997; 97US-0055106P.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Bogdanove AJ, Kim JF, Wei Z, Beer SV;
XX
XX WPI; 2001-327491/34.
XX
DR
XX
PT Polynucleotides encoding hypersensitive response eliciting proteins or
PT polypeptides useful for imparting disease resistance to plants, to
XX
PS Example 9; Col 34; 37pp; English.
XX
CC The sequence represents an Erwinia amylovora disease-specific region E
CC (dspE) mutant polypeptide, which elicits a hypersensitive response in
CC plants. The dspE operon functions as an avirulence (avr) locus and the
CC dspE protein is required for pathogenicity of Erwinia amylovora. DspE is
CC also responsible for causing fireblight in certain plants. The protein
CC and its associated nucleic acid can be applied in a non-infectious form
CC to plants or plant seeds to impart disease resistance, to enhance plant
CC growth, and/or to control insects on plants. Alternatively, plant cells
CC may be transformed with the DNA to form transgenic plants with the same
CC properties. The method can be utilised to treat a wide variety of plants
CC and seeds, including crop plants such as rice, wheat, corn, cabbage,
CC cauliflower, garlic, cucumber, apple, grape, tobacco and sugarcane, and
CC ornamental plants such as Arabidopsis thaliana, petunia, chrysanthemum
XX
SQ Sequence 518 AA;
Query Match 28.8%; Score 2723; DB 4; Length 518;
Best Local Similarity 100.0%; Pred. No. 4.1e-163;
Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 203 GSHHEIKKEEPVGTSTKATTAHADRVETIAQEDDSEFQQLHQRLARERENPPQPKLGVA 262
DB 1 GSHHEIKKEEPVGTSTKATTAHADRVETIAQEDDSEFQQLHQRLARERENPPQPKLGVA 60
QY 263 TPISARFQPKLTAVAESVLEGTDTTQSPKPSMLKSGAGVTPPLATLDKGLQAPDN 322
DB 61 TPISARFQPKLTAVAESVLEGTDTTQSPKPSMLKSGAGVTPPLATLDKGLQAPDN 120
QY 323 PPAINTLLKQTLGKDTQHYLAHAHASSDGSQHLLDNKGHLFDIKSTATSYSVLHNSHPGE 382
DB 121 PPAINTLLKQTLGKDTQHYLAHAHASSDGSQHLLDNKGHLFDIKSTATSYSVLHNSHPGE 180

```

```

QY 383 IKGKLAQAGTGSVSDGKSGKISLGSQTSHNKTMLSQPGEAHRSLLTGTIWOHPAGAARP 442
DB 181 IKGKLAQAGTGSVSDGKSGKISLGSQTSHNKTMLSQPGEAHRSLLTGTIWOHPAGAARP 240
QY 443 QGESIRLHDDKIHILHPELGVWQSADKDTSHQSROADGKLYALKONRTLONLSDNKSSE 502
DB 241 QGESIRLHDDKIHILHPELGVWQSADKDTSHQSROADGKLYALKONRTLONLSDNKSSE 300
QY 503 KLVDKIKSYSDRGQVAILTDTTPGRHKMSIMPSLDASPEHSISLHFADAHQGLLHGK 562
DB 301 KLVDKIKSYSDRGQVAILTDTTPGRHKMSIMPSLDASPEHSISLHFADAHQGLLHGK 360
QY 563 SELEAQSVAIISHGRLLVADSEGLFSAAIIPKQGDGNELXKAMPQHALDEHFHGHQISG 622
DB 361 SELEAQSVAIISHGRLLVADSEGLFSAAIIPKQGDGNELXKAMPQHALDEHFHGHQISG 420
QY 623 FPHDDHGQNALVKNFRQOHACPLGNHDHGFHPCWNLTDALVIDNQLGLHHTNPPEHIL 682
DB 421 FPHDDHGQNALVKNFRQOHACPLGNHDHGFHPCWNLTDALVIDNQLGLHHTNPPEHIL 480
QY 683 DMGHLSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 720
DB 481 DMGHLSLALQEGKLHYFDQLTKGWTGAESDCKQLKKG 518
RESULT 7
AAU04491
ID AAU04491 standard; protein; 507 AA.
XX
AC AAU04491;
XX
DT 23-OCT-2001 (first entry)
XX
DE Erwinia amylovora disease-specific region E (dspE) mutant protein #2.
XX
KW Disease-specific region E; dspE; insect control; disease resistance;
KW hypersensitive response; rice; wheat; corn; cabbage; cauliflower; garlic;
KW cucumber; apple; grape; tobacco; sugarcane; Arabidopsis thaliana; avr;
KW petunia; chrysanthemum; carnation; transgenic plant; avirulence locus;
KW pathogenicity; fireblight; mutant; mutein.
XX
OS Erwinia amylovora.
OS Synthetic.
XX
PN US6228644-B1.
XX
PD 08-MAY-2001.
XX
PF 22-JUL-1998; 98US-00120663.
XX
PR 06-AUG-1997; 97US-0055106P.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Bogdanove AJ, Kim JF, Wei Z, Beer SV;
XX
XX WPI; 2001-327491/34.
XX
DR
XX
PT Polynucleotides encoding hypersensitive response eliciting proteins or
PT polypeptides useful for imparting disease resistance to plants, to
XX
PS Example 9; Col 34; 37pp; English.
XX
CC The sequence represents an Erwinia amylovora disease-specific region E
CC (dspE) mutant polypeptide, which elicits a hypersensitive response in
CC plants. The dspE operon functions as an avirulence (avr) locus and the
CC dspE protein is required for pathogenicity of Erwinia amylovora. DspE is
CC also responsible for causing fireblight in certain plants. The protein
CC and its associated nucleic acid can be applied in a non-infectious form
CC to plants or plant seeds to impart disease resistance, to enhance plant
CC growth, and/or to control insects on plants. Alternatively, plant cells
CC may be transformed with the DNA to form transgenic plants with the same

```

CC properties. The method can be utilised to treat a wide variety of plants  
 CC and seeds, including crop plants such as rice, wheat, corn, cabbage,  
 CC cauliflower, garlic, cucumber, apple, grape, tobacco and sugarcane, and  
 CC ornamental plants such as Arabidopsis thaliana, petunia, chrysanthemum  
 CC and carnation  
 XX  
 SQ Sequence 507 AA;

Query Match 27.1%; Score 2563; DB 4; Length 507;  
 Best Local Similarity 100.0%; Pred. No. 5e-153;  
 Matches 507; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1064 TVLGHQGVLSKNGEINSEFPSPGKALVQSFNNRSGDLSKSLQQAQVHATPPSAESKL 1123  
 Db 1 TVLGHQGVLSKNGEINSEFPSPGKALVQSFNNRSGDLSKSLQQAQVHATPPSAESKL 60  
 QY 1124 QSMGLGHFVSAGVDSHQGEIPLGRQDPNDKTLTKSLILDTVTIGELHELADKAKLV 1183  
 Db 61 QSMGLGHFVSAGVDSHQGEIPLGRQDPNDKTLTKSLILDTVTIGELHELADKAKLV 120  
 QY 1184 SDHKPDADQIKQLRQOFTDLREKRYESNPVKHYTDMGFTHNKALEANDAVKAFINAFKK 1243  
 Db 121 SDHKPDADQIKQLRQOFTDLREKRYESNPVKHYTDMGFTHNKALEANDAVKAFINAFKK 180  
 QY 1244 EHHGVNLTTRTVLSQSGAEIAKLLKNTLLSDGESMSFSRSGGVSTVFPVTLTKKV 1303  
 Db 181 EHHGVNLTTRTVLSQSGAEIAKLLKNTLLSDGESMSFSRSGGVSTVFPVTLTKKV 240  
 QY 1304 PVPVPGAGITLDRAYNLSFRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTTSA 1363  
 Db 241 PVPVPGAGITLDRAYNLSFRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTTSA 300  
 QY 1364 GNASDWLSAKHKISPDLRIGAAVSTLQGTQLQNSLKFKLTDELPGFTIHLTGHTLTPAE 1423  
 Db 301 GNASDWLSAKHKISPDLRIGAAVSTLQGTQLQNSLKFKLTDELPGFTIHLTGHTLTPAE 360  
 QY 1424 LLQKGIHOMKQSKLTFSDVTSANLDIRAGINLNEDGSKPVGTVARVAGLSANLAA 1483  
 Db 361 LLQKGIHOMKQSKLTFSDVTSANLDIRAGINLNEDGSKPVGTVARVAGLSANLAA 420  
 QY 1484 GSRERSTTSGOFGSTTSANNRPTFLNGVGAGANLTAALGVHSSTHEGKPVGIFPAPTS 1543  
 Db 421 GSRERSTTSGOFGSTTSANNRPTFLNGVGAGANLTAALGVHSSTHEGKPVGIFPAPTS 480  
 QY 1544 TNVSAALALDNRSTQSISLELKRAEPPV 1570  
 Db 481 TNVSAALALDNRSTQSISLELKRAEPPV 507

RESULT 8  
 AAY69258  
 ID AAY69258 standard; protein; 201 AA.

XX AC AAY69258;  
 XX DT 30-MAY-2000 (first entry)  
 XX DE Amino acid sequence of the secretion signal of the DspE protein.  
 XX KW Type III secretion system; DspE protein; hairpin secretion system;  
 XX KW effector protein; Avr protein; avirulence protein; agriculture.  
 XX OS Erwinia amylovora.  
 XX PN WO200002996-A2.  
 XX PD 20-JAN-2000.  
 XX PF 09-JUL-1999; 99WO-US015425.  
 XX PR 10-JUL-1998; 98US-0092357P.  
 XX PA (CORR ) CORNELL RES FOUND INC.

XX PI Bauer DW, Beier SV, Bogdanove AJ, Collmer A, Ham JH;  
 XX DR WPI; 2000-182218/16.  
 XX PT New DNA construct encoding type III secretion system, used for  
 XX PT recombinant production of secreted protein and for identifying potential  
 XX PS effector proteins.  
 XX PT Claim 6; Page 14-15; 71pp; English.  
 XX CC The present sequence represents a functional type III secretion system  
 CC from the N-terminal domain of the DspE protein. The DspE protein is  
 CC secreted by a type III secretion system. The DspE secretion signal is  
 CC compatible with the hairpin secretion system of Erwinia amylovora. The  
 CC secretion signal polynucleotide is used to produce the constructs of the  
 CC invention, which also comprise a promoter and a second DNA that encodes a  
 CC protein or polypeptide that can be secreted by the type III secretion  
 CC signal. Host cells containing the constructs are used for recombinant  
 CC production of the polypeptide. The constructs are also used to screen for  
 CC potential effector proteins, e.g. Avr (avirulence) proteins potentially  
 CC useful in agriculture  
 XX SQ Sequence 201 AA;

Query Match 11.0%; Score 1041; DB 3; Length 201;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-57;  
 Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MELKSLGTEHKAHVTAHNPVGHGVALQQGSSSSPQNAASAAEGKNGKMPRIHQ 60  
 Db 1 MELKSLGTEHKAHVTAHNPVGHGVALQQGSSSSPQNAASAAEGKNGKMPRIHQ 60  
 QY 61 STAADGISAAHQKKSFLRCGLTKKFSRSPAPQGPOTTHSKGATLRLDLDGGETQH 120  
 Db 61 STAADGISAAHQKKSFLRCGLTKKFSRSPAPQGPOTTHSKGATLRLDLDGGETQH 120  
 QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPWVGSGGDEKVPVTPQKRHQLNFGQMRQTWLS 180  
 Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPWVGSGGDEKVPVTPQKRHQLNFGQMRQTWLS 180  
 QY 181 KMAHPASANAGDRLQHSPPHI 201  
 Db 181 KMAHPASANAGDRLQHSPPHI 201

RESULT 9  
 ABM70325  
 ID ABM70325 standard; protein; 3041 AA.

XX AC ABM70325;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Photorhabdus luminescens protein sequence #3422.  
 XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX OS Photorhabdus luminescens.  
 XX PN WO200294867-A2.  
 XX PD 28-NOV-2002.  
 XX PF 07-FEB-2002; 2002WO-IB003040.  
 XX PR 07-FEB-2001; 2001FR-00001659.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.







549 TDR---LNNRQGTIASQ--GEDL-HLTAHQADNNQGTQVLAAGNKGKLSLNTQRWLGD-----598  
 369 ATSYSVLHNSHPGEIKGLAAGTGSVS-----VDGKSGKISLGGTQSHKNTWLSQ 420  
 599 -----XGKLLTNGTLTIQAGELQNLHAETQAGQITINADTLSHQSGVMQ 643  
 421 PGEAHSLLTGIWHPAGARPQG-----PSIRLHDDKIHLHPGLWQWQSAKDTHSOLS 476  
 644 WGDLDLSLTRLIDNHSGLTIAGNGLNLKATTVDNRHGNIVAADQGSGLKTLTKDT---LD 700  
 477 ROADGKL---YALK-----DNR-----TLQNL-----496  
 701 NOS-GKLEAGHALQLSATQLDNRSGSIVAAGDSATLTVGKTIQNAHGHLEAOTLTTTSQ 759  
 497 --DNKSSEKLVKIKSVS-----VDQGVQVAI-----LTDTPGRHKMSIMPSLDA 539  
 760 TLDNTQGVLLAQNIDQTTGHPPTNTAGQVIAEDTLTVNSGQLDNTAGLLQAGREMAVDT 819  
 540 SPESHISLSLHFADAHQGLHKGSELEAQAQVAISH-GRLVVADSEGLKFSAAIPKQGDGN 598  
 820 --HGHGLTNTTHADQKAGRLSGGLTLRTGDDIDNTGGMIAADGKTLTSTAL-----N 871  
 599 ELKMKAMPQALDEHFHGHQHSISGFFHDDHGQNLALVNNFRQOHACPLGNDHQFHPGWN 658  
 872 NTQGGIAGNGGLDIH-----SQQLINREGTLQSDAULTD-----906  
 659 LTDALVIDNOLGLHHTNPPEHILD-----MGHL--GSLAL-----Q 693  
 907 -TGQLLDNOG-----QIMGEKTTITSGPLDNRHGHLOGGLAQDITRHAALDNR 956  
 694 EKLHVPDQLTKGWTAESCKQ-----LKKGLDGAAYLLKXGEVVKRLN-----IN 739  
 957 DGKLLSTDTLTLNTHQLDNRHGQVAVGNTVLNVKTQTDNTGGLRGITQTLTNTAHLN 1016  
 740 QSTSSIKHGE-----NVFSLPHVRNKPPEGDALQGLNKDDKAQAMAVGVNKKYALTE 793  
 1017 RETAHTDNGLEAQNLTVAQOVONTQGAALRAANHLQOVSHLENAOGLISAGKQLTVGD 1076  
 794 KGDIRSPOI--KPTQQLERPAQTLREGISGE-----LKDIDHVDHKQLYA 838  
 1077 ATPASLVINNRQTLTAGQA-TINAHGLSGDQLLSQGDMAVLTLEDPH--HTGNTAA 1133  
 839 ---LT-----HGEVPHQPREAWQNGAE-----SSWHKALPOSESKLSLDM 879  
 1134 NGNLTTLKTGNLINDROIKAGQTLHLDAHLNLTNSASGEISAGQTQIQVHDTLNTTGLIDG 1193  
 880 SHEKPIATPED-----QSOHOLKAGWH-----AYAAPERGEPLAVGTSGSQ-- 921  
 1194 GLTHLTANTLNNMGTRIGYDQLALQGTTLNNTAONGKAAVIAARDRLDITGTGLNNSDH 1253  
 922 ---TVFNRLMOGVKGVIPGSGGLTVKLSAQGTGMTGAERKVSSEKFSERIRAYAFNPTM 977  
 1254 AQIYSVGDHLTGRLDNTLTATQARVLNHAATIEAGRNLIQADRNNTINAGLVTVQV 1313  
 978 STPR-PINKAAYA---TOHGHW-----GREGLKELYEQ-----GA 1009  
 1314 ETEKSPHDAVLSGQTTTRYDSQVDTSRHNKYGVHDAIMPDPGNSNDFEYQYTRTVRET 1373  
 1010 LKOLD-----AHNVRNAPQPDLOSKLETLIDGEGHAELLNDKMRFRDEQSAT--- 1060  
 1374 QVKOSDPKILAGNITLNSAQVTNHS-QIVACGTLGGEI-----GELHNIATQGE 1424  
 1061 RSVTVLG-----QHOGVLKNGEINSEFKPSPG-----KALVQSFNVNRS 1100  
 1425 RIITDTGSGTHWYAKKRLKPRFGTKTSQKSRSRVAPAVIETIDLKTLAWQAHRPQ 1494  
 1101 QDLKSLSLOQA--VHATPPSRESKLSQMLGHFVSAGVDMHSHQGEIP-----1145  
 1485 GTDITITDROGTQHAAPTA-----VTFVNGINNQLVLPQGFELSLPPE 1532  
 1146 -LGRQDNDKTLTKSL-----ILDVTITIG-ELHELAD-----KAKLVSD-----1185  
 1533 VKGQTVDPVIRVVPDTRLPNNSLTVQPGSDSHVLETDPKFTQYKQWLGSQWYREQLT 1592

1186 -----HKPAD---QIKOLROQFDTLREKRYESNPVKHYITDMGFTTHKALEANYDAVKA 1236  
 1593 HDPALVHKRLGDFEQRLVRDQITQLTQGRYLS-----GYND---EAQKALMD 1640  
 1237 FINAFKKEHI--GVN-----LTRTVLESQGSABIAK-----1266  
 1641 AGVAFKQOQLTPGVALSAPQWALLTSDIWLNTQVTUPLDGTTEVTVTVQVVARVROGD 1700  
 1267 -----KLXN-----TLLSLDSGESMSFSRSYGGVSTVFVPT 1298  
 1701 LRSDGALLAGNTVALNNOGDIINSGTISGRDVTQLTANNLTNSGFI--GGKVDLAAQQT 1758  
 1299 LSXK-----VPVPVTPAGAIT-----LDRAVNLFSFSTSGGLNVSF-----1334  
 1759 LTRNGQIQGDDRVTLKGRDITSASTVRGDEANRWLDRPAGIYVQNDKGTLSLSAINNVQ 1818  
 1335 -----GRDGVGSGNINVATGHDV-MPYMTGKKTISAGNASDW--LSAKHKISPDLRIGAAV 1386  
 1819 LTASDVKNAGKOGHTETAGHNLTLDAALSTRTEQG--DWKDNTRH-LTQQODIGSQI 1874  
 1387 SGTQGTQLQNSLKFCLTEDELPGFIHGLTHG-----TLTPAELLQKGIHEHQMGQSK-----1438  
 1875 TGTGEVTLQAGQDLNATAAHVNAGQHLTAQAGNSLTLTCTAGSDLVEHS-KQTSKGWLS 1933  
 1439 -----LTFSDVTSANLDIRAGINLNEGSKPENG-VARVSAGLSASAN 1480  
 1934 KSSVETHDEVDHQRALSTFTPSGD---KVNLAQKDLNIRGNSVAGTQDVSLNAGHQUTVT 1990  
 1481 LAAGRSERS-----TTSQ-----QFGSTTSASNRRPTFL 1509  
 1991 TAAESHDETHLROBKXGSLMGTCGIGFTLKASQKVTYDSDRQLSKGSTVSSQGNVTLN 2050  
 1510 NG-----VGAGANLTAALGVAHASTH-EGKPVGIFPAPTSTNVSAAL 1550  
 2051 AGQLRVHSGEVIAGKDLTLTGQVDITSANRHHHTTTKTEQKQSGLTVALS-----GAAG 2106  
 1551 ALDNRTSQISLSLEKRAEP--VTSNDISELTSLGKHFKDSATTMKLAALKELDD---1603  
 2107 GAVNSAVQTARAAARTESDPRVKALONTQAALSGV-----QAAQAGRLAAQSGDDKGN 2160  
 1604 -----AKPAEOLH--ILQO--HPSAKD-----VVGDER-----1627  
 2161 NLAVLSYSGRQSRSEQHRQTTQOGSHLTAGDNLITAKGDDKAGSQNGDIRIQSGQ 2220  
 1628 YEAVRNKLKLVIRQ-QAADSHMELGSAHST-----TYNNLS-----1664  
 2221 LQAGKDLQNAHRDILQLSSQNTETQTKNSSHGSGALGVGLTVGPGGTGLNISANVSRGN 2280  
 1665 -RINNDGIVELLKHFDPAALPASSAKELGEMMNDPALKDIIKQLOSTPSSASVSMELK 1723  
 2281 GRENGGV-----SHTNTTLOA--GQTVGLNSGRDITLK-----GAQVSGEQLTAEVK 2326  
 1724 DGL-----REQTEKAILDGKVGREEVGV--LFQDRNLRVKSYSVSQSVSKS 1768  
 2327 RHLTSLSEQDSQRYDSQQONASAGVSATVGLTINGTASLNASRNKLSHNVDSVQEQTLGF 2386  
 1769 EGFNTPALILG-----TNSAAMSERNITGTINFKYQDQNTPRFTLE---GGI 1815  
 2387 AGQGGYQVNVGDHFTQLDGAVIASQADKANNTLTNTGLGFK---DKNQADFTVEQQSAGI 2443  
 1816 AQANPQVASALTDLKKEGL 1834  
 2444 SLQGTAGQVLNLAUVNGL 2462

RESULT 11  
 ABM67717  
 ID ABM67717 standard; protein; 2951 AA.  
 XX  
 AC  
 XX  
 XX  
 DT 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #814.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

XX KW detection; food; gene expression; plant; animal; microorganism; toxin;

XX KW antibiotic; biopesticide; virulence factor; disease model; plague;

XX KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN WO200294867-A2.

XX XX 28-NOV-2002.

XX PD 07-FEB-2002; 2002WO-IB003040.

XX PF 07-FEB-2001; 2001FR-00001659.

XX PR (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX FA Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

XX PI Buchrieser C;

XX PI WPI; 2003-148459/14.

XX DR Genomic sequence of Photorhabdus luminescens and encoded polypeptides,

XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX XX Claim 2; SEQ ID NO 814; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded

XX CC proteins from Photorhabdus luminescens. The isolated sequences are

XX CC sources of probes and primers for detecting the genome of P. luminescens

XX CC and related species; to study polymorphisms; for gene analysis and for

XX CC detection/amplification of the genes. Antibodies (Ab) raised against the

XX CC polypeptides encoded by the genes are used for detection/identification

XX CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that

XX CC carry a gene-containing vector are used to select compounds that

XX CC modulate, regulate, induce or inhibit expression of the genes in plants,

XX CC animals or microorganisms other than P. luminescens and are able to alter

XX CC response or sensitivity to toxins and antibiotics produced by P.

XX CC luminescens. Cells transformed to express the genes are useful for

XX CC recombinant production of the proteins, particularly toxins and

XX CC antibacterials useful as insecticides, bactericides and fungicides. The

XX CC genes, proteins, vectors containing the genes and Ab are also useful

XX CC therapeutically (to treat microbial infection by bacteria or fungi that

XX CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as

XX CC biopesticides. Other uses of the genes and the proteins are as virulence

XX CC factors and for identifying targets of human diseases for which P.

XX CC luminescens is a model (particularly plague and whooping cough). This

XX CC sequence represents one of the isolated P. luminescens proteins

XX XX Sequence 2951 AA;

Query Match 3.1%; Score 295; DB 6; Length 2951;

Best Local Similarity 19.3%; Pred. No. 1.1e-06;

Matches 427; Conservative 299; Mismatches 769; Indels 720; Gaps 110;

QY 2 ELKSLGTEHKAHVTAHNPVGHVA-----LQQSSSSSSPONAASLAEGKNRGM 54

DB 308 KIRLIGFETGVGHNAIGNIGASAGEVHTADGWVENRGTLSRDLTQLTSSADVTNGKL 367

QY 55 PRIQHPSTAADGISAAHQKKSFSLRGLCTGKFKPSRAPOG---QPGTTHSKGATLRDL 110

DB 368 -----LSQSAVNLOTKAL--RNGRVEARGDTTVNAGTIHSHSDSVWAA 410

QY 111 LARDGDT-----QH-EAAPDAARLTRSGGVKRRNMD-----MAG 146

DB 411 GLDDNGNTTRPGLSLTAQHVAQKGNLAADTLAVHSQRIDLSQTAASQIQLTASQAG 470

QY 147 RPWVKGSSEKDYPTQQRHQLNFGQMRQTMLSKMAHPASANAGRLQHPHPHPSGH 206

DB 471 ISTARATVNADRL-TAKTPGFQFNNDGQ---LVAKIEHLTT-----PDLNSNOQG 515

QY 207 EIKPEPVGSTKAT-----TAHADRVIAQEDDDSEFQOLHQRLARER 250

DB 516 KINOTGTGELTLHTRTLNLRREGTVFNQOKLTLTDLRLNNRQGTIASQGEDLH----- 567

QY 251 ENPPQPKLVGATPISARFQPKLTAVABSVEGT---DTTQSPKPKPSMLKSGAGVTP 306

DB 568 -----LTAHADNNGQTVOLAGQGRSLNLTORWLGDGKLLTN 605

QY 307 LAVTLDKGKLOLAPNPPA-----LNTL-----LKQTLGKDTQHYLAHHAASSDGSOHL 355

DB 606 GALTQAGELQNAOETQAGQITVNADTLSHQGVMOQOGKDT---LSLTKRLDNQRT 662

QY 356 LDNGHLFDIKSTATSYVLHNSHPGEIKKLAQAGTGV-----SVDGSGKI----- 404

DB 663 LAGNGL-NKAT-----TVDNRH-----GNIVAAEKGSLTAVKDTLDNOSGRVLSGGQ 711

QY 405 -SLGSGTQSHNKTMLSQPGEAHRSLITGIWHPAGAARPOGE-----SIRLHDDDKIHILH 458

DB 712 LTLRTGVDNTGGIIADG---KTLTSS-----AELTNQOIAGNGGLDIHQO--LIN 761

QY 459 PELGWQSGADKTHSQLSRQADGKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQR-- 516

DB 762 RE-GTLQASD-----ALTLDTDGQLL---DNQOQHILGEGKTT-----ITSGPLDNRHG 806

QY 517 ---GOVAILLDTFGRHKMSIMPSIDASPESHISLSLHFADAHQ----- 556

DB 807 HLOGGQLAI--DT--RH-----AALDNRDGKLLSTDTLTNLTHQLDNHRHGQVAVGDTTL 857

QY 557 -----GLHKGSE-----LEAQVAISHGLVVAADSEGLF 587

DB 858 NVDTQTDNTGGLIRGGQQLTLNTAHLINRETKQTDNGLEAQLNLVNAQO--VDNTQGAIR 915

QY 588 SAALPK-----QG--DGNELKWA--MPQHALDEHFG-----HDHQIS 621

DB 916 AANHLKANIQTLLNNTQGLVNSAGKQLILKSEARPHLAINNRQGTIAGEHATINAHLS 975

QY 622 G-----FFHDDHCOLNALVKNFRQHQHACPLGNDHQPFG-----WNLTD 661

DB 976 GDGQLLSQGDMAVTLTDFHTHTGNTAANGNLTLTKTGNLLNDRQIKAGRALHLTHNLN 1035

QY 662 ALVIDNOLGLHHTNPEPEHILD-----MHLGLSLAQE--GKLHYFDQLTKGTGA 710

DB 1036 S--ASGISAGQIQIAHDVLSNTGLIDGGLTHLTANTLNTGTGRI-YGDQAL-QTGT 1091

QY 711 ESDCKLKKGLDGAAYLLKDGVEKRLNINQSTSIKHTENVFSLPHVRNKPPEGDALQG 770

DB 1092 LNNTAQ-----DGKAAVIAARD--RLDIGTATLNNQHAQ-IYSV-----GDMRIG 1134

QY 771 LNKDDKAQAMAVIGVKNKYLALTEKGDIRSFIQKPGTQOOLERPAOTLSREGISGELKDIHV 830

DB 1135 GQLDN-----TLTASGQ-----AHELANNHATIE-----AGRHLKINA 1167

QY 831 DHKQNLVYALTHEGEVHFQPREAWQNGAESSSWHKIALPQSESKL--KSLDMSHBK--P 885

DB 1168 DQINN---TNAGLVITQVVE-----TEKSOHDAVLSGQTRYDWSQVDTSHRNKKYKH 1217

QY 886 IATFEGSQHQLKAGGHAYAAFERGLAVGTSGSQTV-FNRLMQGVKGKVPISGLITVK 944

DB 1218 DAIMPDS-----RSNDFEY-----QVTRTVKETQVKQSDPGKILAGNITL- 1260

QY 945 LSAQTGGMTCABGRKVSSEKSEIRAYAFNPTMSTPRPIKAAAYATGHQWQREGKLPY 1004

DB 1261 -----NSAQVTNHDQ-----I 1272

QY 1005 EMOGALIKQI--DAHNVRHNAPODLOSKLETLDLGE--HGAELLNDMKRFRDELEQSATR 1061

DB 1273 VASGALDGEIGELHNTA-----TQGERITTDKGQTHWYAKTKHKPF----- 1315

QY 1062 SVTVLGOHQGVKLSNGEINSEFKPSFG-----KALVQSFNVNRSGQDLKSLSQAA--VH 1113

DB 1316 -----GFGTKTKSQKRSRNYDPAPVIEITDLKTLAWQENTRPGQTDITITDQTQI 1369

QY	1114	ATPPAESKLSQMLGHFVSAGVDMSHQKEIP-----LGRQDPNDKTAITKSLR-----ILD	1166	KW	detection; food; gene expression; plant; animal; microorganism; toxin;
Db	1370	SAFTA-----VTPVNGIKNPVLVPETVKGTVDVIRVVTPDRLPNNSLY	1417	KW	antibiotic; biopesticide; virulence factor; disease model; plague;
QY	1167	TVTIGELBELADKAKLVSHKPDADQIKQ-----LRQF-----DTLEKRY	1208	OS	whooping cough.
Db	1418	TVQPG-----SDSHYLV-ETDPKFTQYKQWLGSDYMRQQLTHDPALVHKRLGDFGEQRL	1471	XX	Photorhabdus luminescens.
QY	1209	ESNPVKHYTDM-----GFTINKALEANDYAVKAFINAFKEHH-----GVN	1249	XX	WO200294867-A2.
Db	1472	VRDQITQLTGQRYLPQYNNND-----EAQFKALMDAGVAFKQOQLTPGVALSQAQMALITSD	1528	XX	28-NOV-2002.
QY	1250	---LTPRTVLESQGSALA-----KKLKNLTLSLD-----S	1277	XX	07-FEB-2002; 2002WO-IB003040.
Db	1529	IIWLNTQVTLPGTTEVVTPQVYARVRQGDLSGALLAGNTVALNSQGDITNSGTMS	1588	XX	07-FEB-2001; 2001FR-00001659.
QY	1278	GSMS-----FSRSYGGGVSTVFPILSKK-----VPPVPIGAGIT-----	1314	PA	(INSP ) INST PASTEUR.
Db	1589	GRDVTQLTANNLTNSGFIR--GGKVDLAAQQTLTNRGGQIQGDDRVTLKGRDITASTVR	1646	PA	(CNRS ) CNRS CENT NAT RECH SCI.
QY	1315	-----LDRAYNLSPRSTSGGLNVGF-----GRDGGVSGNIMVATGHV-MPYM	1356	PI	Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
Db	1647	GDANRWLDRPAGIYQNDKGTLSLAINNVQLTASDVKNAGDKGHTETAGHNLTLDAL	1706	PI	Buchrieser C;
QY	1357	TGKKTSGNASDW--LSAKHKISPLDRIGAASVGTLOQTQNSLKFKLTEDEL-----	1407	XX	WPI; 2003-148459/14.
Db	1707	STRTTEQG--DWGKDNTRLTQQDQIGSQITGAGEVTLQAGQDLNATAAHVNAVQQLTA	1763	XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
QY	1408	PGFIHGLTHTLTPAELLQKGTIEH--OMQGSKLTFSDVTSANLDRAGINLNEGSKP	1464	CC	useful e.g. as therapeutic antimicrobials and agricultural pesticides.
Db	1764	QAGNLTLTGTAS--SDL-----VEHSKQTSKGLWSKSVETHDVDRQALSTFTFGDK-	1817	CC	Claim 2; SEQ ID NO 551; 1205pp; French.
QY	1465	NGVTARVSAGLSASANLAAGSRSTTSQ-QFGSTTSASNRRTEL-----NGVGAGANL	1518	CC	The invention relates to the isolation of genes and their encoded
Db	1818	--VTLQAGKDLNIRGNSVAGTQDVSLNAGHQLTVTTAAEAHDETHLRQEKKSGLMGTSGM	1875	CC	proteins from Photorhabdus luminescens. The isolated sequences are
QY	1519	TAALGVA-----HSSSTHEGPVGIFPAFTSTNVSAALADNRITSQSISLBKRAFPV	1570	CC	sources of probes and primers for detecting the genome of P. luminescens
Db	1876	GFTVKGASQKVTTSDSLSKSGSVSGSVSGNVTLNAGE-----QLRVHSGSEVI	1923	CC	and related species; to study polymorphisms; for gene analysis and for
QY	1571	TSNDIS-----ELTSLTGHFQKDSATTMLAALKELDDAKPABQLHLOOHSAKDVVG	1624	CC	detection/amplification of the genes. Antibodies (Ab) raised against the
Db	1924	AGKDLTLTGQVVDTSIAENRHHHTTKTEQKQSLTVALSGAAGAVNSAVQTAARTES	1983	CC	polypeptides encoded by the genes are used for detection/identification
QY	1625	DERVEAVRNKLKVIROQAADSHSMELGSHSTVYNLNRIN-----NDGIVELLHK--	1677	CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
Db	1984	DPRVKALONTKAALSQVQAAGLAEAGQSDDKGNLNLGAVLSYGRQSRSEQQHROT	2043	CC	carry a gene-containing vector are used to select compounds that
QY	1678	-----HFDA-----ALPASSAKRLGEMMNDPALKDIIKQLQSTFFSSASVSMELKGLRE	1728	CC	modulate, regulate, induce or inhibit expression of the genes in plants,
Db	2044	TQOGSHLTAGDNLITITAKGDKGASGQNGDIRIQG--SQLQAGKDLQNLNARDIQLSSSQ	2101	CC	animals or microorganisms other than P. luminescens and are able to alter
QY	1729	QTEKAILDGKVGREVGVLQDRNRLRVKSVSVSQSVKSEGENTPALLIGTNSAAMS	1788	CC	response or sensitivity to toxins and antibiotics produced by P.
Db	2102	NTEQT-----TGKNSHRSGLMGLTVGPG-----GTGLNVSANV	2136	CC	luminescens. Cells transformed to express the genes are useful for
QY	1789	ERNIGTIN-----FKYQD--QNTPRRFTLEGGIAQANPQVASALTDLKK	1831	CC	recombinant production of the proteins, particularly toxins and
Db	2137	SRNGRENGVSHVHTNTLQAGQSVGLNSGRDITLKG--AQVSGEQLTA--DVKR	2187	CC	antibacterials useful as insecticides, bactericides and fungicides. The
RESULT 12				CC	genes, proteins, vectors containing the genes and Ab are also useful
ABM67454				CC	therapeutically (to treat microbial infection by bacteria or fungi that
ID	ABM67454			CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX	ABM67454			CC	biopesticides. Other uses of the genes and the proteins are as virulence
AC	ABM67454			CC	factors and for identifying targets of human diseases for which P.
XX	20-NOV-2003 (first entry)			CC	luminescens is a model (particularly plague and whooping cough). This
DT	Photorhabdus luminescens protein sequence #551.			CC	sequence represents one of the isolated P. luminescens proteins
DE	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;			XX	
XX				SQ	Sequence 4560 AA;

Query Match 3.0%; Score 280.5; DB 6; Length 4560;  
Best Local Similarity 19.9%; Pred. No. 1.7e-07;  
Matches 440; Conservative 247; Mismatches 786; Indels 735; Gaps 113;  
QY 29 QCGSSSSSPQNAASLAAGCKNEGKMPRIHOPSTAAADGISAHQKKSFLRGLGPKKF 88  
Db 1813 QQGRVITADKOLNVTSTGAVDNTSGKIVSQHQQLTMNTG-----ELNNTSGLLQSKTT 1864  
QY 89 SRGAPOQOPQTHSKGATLRDLARDGGETQHEAAPD--AARLTRSGGVK--RRNMDDM 144  
Db 1865 LYLNTHGKQLNTQSG---DDLGRSDSLTLEAGEIDNTAGKIDSQGETKLTSLNNT 1921  
QY 145 AGRPMVKG-----GSGEDKVFTQQR----- 165  
Db 1922 DGKILSKGKADLTQAINNORGLTQSASLKLDTQQOKLTNTDTSKINGLALDLSLKIT 1981  
QY 166 -HQINFEQMRQOTWLSKWAHPASANAGDRLQSPHPHGGHHEIKERPVGSTSKATTAHA 224  
Db 1982 GELINDFGYR-----GDE-----THI-NSHQQTLSNLTGTIVSKNKL 2020  
QY 225 DRVEIAQ-----EDDDSEFOQLHQORLAREFNPFPKLGVTATPISARFQKLTAVAES 279

```

Db 2021 DSGELDSTGGLIKSDGNMTIDTHGKLLTTAKSGDTR-----GVIS-----EGTMTLTADDE 2070
Qy 280 V-----LEGDTTQSPKQPMKSGAGVTPLAWLTKGKLQADPNPPALNTLLKOT 333
Db 2071 IDNQDFIKGTGT---IVTGGELKNQGG---TLISEKSGLTISVQNTDSGGLL-QS 2121
Qy 334 LQK---DTQ-HYLAHASSD---SQHULLDNKGHLFDIKSTATSYSV-----LHNSHP 380
Db 2122 AGKLTLDTHGSLTNKNSGDRGIRSQDDMLITSGDLHNQAGTINRKTATVNNLQNTIV 2181
Qy 381 G-EIK-----GKLAQAG-----TGS----- 394
Db 2182 GTQQLKLTQAPNNQGGKVHSDGNLNLNTQGRNLANTGDKRSGSFSAQGDLTLDIGELN 2241
Qy 395 -----VSDGKS-----GKISLCSGTQSHNKTMLSPG-----EAHRSI 428
Db 2242 NDASFIADGKTITSTLTNNKGLIAGNSGLEIHSQTLTNSGSLKSNANTVNIQNL 2301
Qy 429 LTGIWQHAPGAARPOGESIRLHDDKHILHPGLVWQSDADKTHSGLSRQADGKLYALKD 488
Db 2302 LDNQHGRIIGDNTTTSGLKNNQHGHIOGKKLTI-NTQADTDNQ-----DGKLLS-AD 2354
Qy 489 NRTLQNL-S-DNKSSEKLVKISYVDORGVAAILTD-----TPGRHKMSIMPS 536
Db 2355 TNNLTLOLDR-----RGVKAIGDATINAPKQNTNNTGGLIHSDQQLT 2398
Qy 537 L-----DASPESH-ISLSLHFADAHQGLLHGKSELEA---QSVASHG----- 575
Db 2399 LKTABLINRETNHPDQGTAEADLIIEAQIDNTQGTLOGANPLQAIINQSLKNEQGLISG 2458
Qy 576 --RLVVADSEKLSAATPKQ--DGNE-LKKAMPQHALDEHFGHDHQISGFFPHDDHQ 630
Db 2459 GKQLTIEDTKQL--TVNNQGTTLTGSEKVIKA---NAL-----SGDGQIL-----SQGD 2504
Qy 631 LNALVKNFRQOQHACPLNGDHQFHPGNWLTDAVLVDNQGLHHTNPEPEHILDMGHLSL 690
Db 2505 IEVKLKQDFHN-----TGNIAADGKLSUETD-----GNIIND 2536
Qy 691 ALQEGKLYFD--QUTKGWTAESCKQ-----LKKGLDGAAYLLKDGVEVKRLINQSTSS 744
Db 2537 STMKAEQAYLEAQLNTQT-AEIRAKQTEVNVNRLTNTG--LIDGELTHLTANKVLDN 2593
Qy 745 IKHGTENVFSLPHVRNKEPEGD-----ALQGLNKDQKAAQAMAVIGNKYALTEKGD 797
Db 2594 T--GTGRY-----GDQIALKSGTLNNTAKDGKA---AVIAARDL----- 2629
Qy 798 RSFOIKPQTQQLERPAQ--TLISREGISGELKDIHVHDKONLYALTHEGEVHPQPREAQW 855
Db 2630 ---DIGTGLNRRNHAQIYSVDMRIGGQL-----DENLAATSQAGRL----- 2669
Qy 856 GAESSSWHKLALPOSES-KLXSLDMSHEHKPIAT----FEDGSOHLKAGG-----WHA 904
Db 2670 -----SNHAATIEAGSLKIDAAIINKNTNGLVTDIVETEKSORHEAVLSGRTARYDWSQ 2724
Qy 905 YAAERGPPLAY-----GTSSQTVF-----NRLMOGVKQKVPISGSLTVKLSAQ 948
Db 2725 VNTSRNKYGVHRAVMPDPSGSEEFYRYRTTINETQIKESDPGKILSGNNTINSAML 2784
Qy 949 T-----GGMGTGAE-----GRKVSXKFSERIRAYA-----FNPTMST----- 979
Db 2785 TNDSDQIVAGIGLGDIDELHNNATKGERIITDEGSEIRWYAKKKRRFRGKTSSQGDW 2844
Qy 980 ----PRPIKNAAYATQHCWQGREGLKPLYEMQGALIKQLDADHNVNRHNAQPQDLOSKLETL 1035
Db 2845 DHYNPAPITETIDLKALAWQ--NTRP--NSTGITIDDRQTSRVQSTPTGINLTSRM--A 2898
Qy 1036 DLGE-----HGALLNDMKRRFDELEQSATRSVTVLQGHQGVKNSGEBINSEFKPSP 1087
Db 2899 EAGEVFTFNATINATLPN-----RDIPTDRPLLSPT--GQOTEQLLSSGAV-----ALP 2946
Qy 1088 GK-ALVQSFNVNRSGOD-----LSKS-----LQQAHVHATPPSAESKLQ 1124
Db 2947 SRDALTEHPTLSGQQTQEQVLSSSAVVLPNRDALTEHPVLPVQQTTERVLPP----- 2999

```

RESULT 13

AAV75097

ID AAV75097 standard; protein; 2514 AA.

XX AAV75097;

XX AAV75097;

XX AAV75097;

DT 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 564 protein sequence SEQ ID NO:1668.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX WO9957280-A2.  
 XX  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US009346.  
 XX  
 PR 01-MAY-1998; 98US-0083758P.  
 PR 31-JUL-1998; 98US-0094869P.  
 PR 02-SEP-1998; 98US-0098994P.  
 PR 03-SEP-1998; 98US-0099062P.  
 PR 09-OCT-1998; 98US-0103749P.  
 PR 09-OCT-1998; 98US-0103794P.  
 PR 09-OCT-1998; 98US-0103796P.  
 PR 25-FEB-1999; 99US-0121528P.  
 XX  
 PA (CHIRON) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter JC;  
 XX  
 DR WPI; 2000-062150/05.  
 DR N-PSDB; AA253859.  
 XX  
 XX Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics.  
 PT  
 XX  
 PS Claim 2; Page 854-855; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA254253 to AA2575941  
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of the  
 CC invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the manufacture  
 CC of medicaments for treating or preventing infection due to *Neisseria*  
 CC bacteria (e.g. meningitis and septicemia), to detect the presence of  
 CC *Neisseria* bacteria, or to raise antibodies. They may also be used to  
 CC screen for agonists or antagonists, which may themselves have use as  
 CC antibacterial agents. The polynucleotides of the invention may also be  
 CC used in gene therapy protocols  
 XX  
 SQ Sequence 2514 AA;  
 Query Match 2.9%; Score 275; DB 3; Length 2514;  
 Best Local Similarity 18.8%; Pred. No. 1.5e-07;  
 Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;  
 QY 7 GTEH-KAAVHTAAHNPV-----GHGVALQQGSSS-----SSPQNAAS 43  
 Db 37 GSAHVKSVPFTTAPVCRNIFSFLLGSLCLAVGTANTAFADGIADKAAKPTQAT 96  
 QY 44 LAAGCKNRGKMPRIHQPTAAGDISAAHQKKSFSLRGCLGKTKKFSRSPAGQPG----- 98  
 Db 97 ILQTGNG--IPQVNIQPTSGAGVSNQYQAFDVGNGAILNN--SRSTNTQTQLGGMIQ 151  
 QY 99 -----TTHSK-----GATRLDLLRDDGTOHEAAAPDAARLTR 132  
 Db 152 NPWLARGEARVVVQINSSSHSQMNGYIEVGGRBAEVVIANPAGIANGGGFINASRTL 211  
 QY 133 SGGVKRRNMDMAGRPVKG-----GSGED-----KVPTQQRHQLNFGQMRQTMLS 180  
 Db 212 TTGQPOYQAGLISGFKIRQGNVVIAGHGLDARDTDFRLLSYHSHKIDAPVWGQVVRVAG 271  
 QY 181 KWAHPASAGDRLOHSPPHIPGSHHEIKFEPVGSSTKATTAHADRVEIAQEDDDSEFQ 240  
 Db 272 QNDVVATGNA-----HSP-----ILNNAANTSNNTANGTHIPLFAIDTG----- 312  
 QY 241 LHQQLARERENFPQPKLVATPISARFQPKLTAVASVLEGTDTTQSPKPKQSMKGS 300

Db 313 -----KLG-----GMVANKITLISTA-----EQAGIRNOGQJFAS 342  
 QY 301 GAGVTPPLAVTDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHAS--SDGSQHLILDN 358  
 Db 343 SGN-----VAIDANGRLVNSGTMAAN-----AKDTNTAEHKVNIRSQ-----VEN 385  
 QY 359 KG-----HLFDIKSTATSYS-----VLNHSHPGEIKKLAQAAGTGSVDVKGSKI 404  
 Db 386 SGTAVSQOQGTQIHSQSQIQTGTLSSGEILHNS-----GSLKNETSGTI-----EAARL 435  
 QY 405 SLGSGTQSHKNTMLSQBGEAHRSLLTIGIWHCPAGAAPQGESIRLHDDKTHILHPELGV 464  
 Db 436 AIDTDT-LNNQGLSOTG-----SQUHI----- 458  
 QY 465 QSADKTHSLSRQADGKLYALKDNRTLQNLSD-----NKSSEKLVDKIKSYSDVDRQG 518  
 Db 459 -----DAQKX-----DNRGMLQDTAPTASNGSSNQIGN-----SYNASPHSS 498  
 QY 519 VAILTPTPRHKMSIMPSLDASPEHSISLHFAHQAQGLLHGKSELEASVALSHGRLV 578  
 Db 499 TTTPTTATGTGTATVISINITAPT-----FAD-----GTIRTHGALDNGSGSIANGQTD 547  
 QY 579 VADSEKLFSAATPKOGDGNELKMKAMPOHALDEHFGHDHQISGFHDDHGHQNALVKN- 637  
 Db 548 VSAQOG-LNNAG--QIDIHQLNAG--SAFDNHNG--TIISDAVHIQAGSLNNQNGNI 598  
 QY 638 NFRQOQHACPLGNDHQFHPGWNLTALVID-NQLGLHHTNPEPEHILDMHGLSLALQEGK 696  
 Db 599 TTRQQ-----LEIETDQLDNAHGLKLSAEIADLAVSGSLNNQNGE 638  
 QY 697 LHVFDQLTKWTGAESDCKQLKGLDGAAYLLKDGVEKRLNINQSTSSIKHGHTENVSILP 756  
 Db 639 IATNQQL-----IHDGQOSTAVIDNTNTINGTIOGRDVAIOAK 675  
 QY 757 HVNKEPPEGALQGLNK-----DDKAQAMAVIGNVKYLALTEKGDIRSFOIKPQTQOLE 810  
 Db 676 SLSN-----NGLTADNKLDTALQODFFVERNIIVAGNE-LSLSRGSLSKNSHTLQAGKRI 730  
 QY 811 RPAQTLREG-----ISGELKDIHVHDKQNLALYALTHEGEVFPHPREAWQNGAESSSWHKLA 866  
 Db 731 IKANNLDNAQGNITQSGTTDITGQHN-----LTNRGLIDGQ----- 767  
 QY 867 LPQSESKLSLDMSH-----EHKPIATFEDGSQHLKAGGWHAAAPERGPLAYGTS 918  
 Db 768 ----QTKIQAGQMNIGTGRIYGNIAATAALDNLQDENGTTGA--ATAARENLMIGI-- 819  
 QY 919 GSQTVFNRLMQGVKVI PGSLTVKLSAOTGGMTGAEGRKVKSSFSEIRIAYAFNPTMS 978  
 Db 820 -----QLNNRENSLIYSGNDMAVGGLDNTNGAQTGAQR----- 853  
 QY 979 TPRPIKNAAYATQHGMOGREGKLPYEMOGALIKQL-----DAHNVRHNAP 1024  
 Db 854 ----IHNAGATIEAAGKMWLGVEKLNHTNBEHLKTOLVETGREHIVDVEAFGRHELLREGT 909  
 QY 1025 QPDL-----QSKLETLDLGEH-----GAELLNDMK 1049  
 Db 910 QHELGSVYNDESDEHLRTPDGAHENWHKYDEYKVTQKTQVOTAPAKIISGNDLTIDGK 969  
 QY 1050 R-PRDELEQSATRSVTVLGOHQ-----VLKSNGEINSEFKPS-PGKALVQSFN 1096  
 Db 970 EVFNTDSQIITAGGNIIVQTEKDGHLNBEQTGEEKVFSNGKLHSYWEKHKGR---DSTG 1026  
 QY 1097 VNRSODLSKSLQOAVHATPPSAESKLSQMLGHFVSAGVDMSHQKGIPIGRQRDPNDKT 1156  
 Db 1027 HSEQNYTLPEITRNI SILGSFAYESHKALSHHAPSQGTLPQSGN---ISLPTYNSFT 1083  
 QY 1157 ALTKSRL-ILDTVTIGELHELADKAK-----LVSDHKPDA---DQIKQLRQQTTLREKR 1207  
 Db 1084 PLPSSSYIINPVNKGYLIVETDPRFANYROWLGSYMLDSLUKLDNNLHKLGGLGYEQR 1143  
 QY 1208 YESNPVXHYTDMGTGTHNKALEFANYDVAKFNAFKKEHHGVNLTATRTVLESQGSAAELAK 1267  
 Db 1144 LINEQIAELT--GHRRLDGYQNDDEEQKALMDNGATAARSNNLSVGIAL-----SAEQVAQ 1197

QY 1268 LKNTLLSDGESMSFSRSGYGVSTVFPTLSKKVPVPIPGAGITLDRAYNLSFGRS 1327  
 Db 1198 LTSIWLWVQKE-----VKLPDGGTQTVLPQVYRVXNGDIDKGAL-----LSGNTQ 1247  
 QY 1328 GGLNVSGRGGVSGNIMVATGHDVPMYTKG---KTSA-----GNASDWLSAKHKIS 1377  
 Db 1248 INVSSLKNSGTIAGRNALINTDLDNIGGRIHAQKSAVATATQDINNIGGMLSAEQTL 1307  
 QY 1378 PDLRIGAAV-SGTLOGTQLNSLKFTEDELPG-FIHGLTHGTLT-----PAELLQKIE 1430  
 Db 1308 --LNAGNNINSQSTASSQNTQSGSTYLDRWAGIYITGKEGVLAAGAKGDINIAGQIS 1365  
 QY 1431 HQMKQSGKLTFSVDTSANLNDL-----RAGINLNDGSKPNQVTVARVSGAGLSASA----- 1479  
 Db 1366 NQSQG-QTRLOAGRDINDLTQVTSKQATHFDADNVIHVRSTNEVGSSIQTKGDVILLS 1424  
 QY 1480 --NLAAGRSRSTSGFGSTTSASNNRPTPLNGVAGANLTAALGVVAHSSTHGKPVGI 1537  
 Db 1425 GNNLNKAAAEVSSANGTL--AVSAKND---IN-ISAGINTT---HVDASAKHTGRSGG- 1473  
 QY 1538 PPAFTSNVSAALADNRSTQSISLELKRAPVTSNDISELTSTL-----GKHFK 1587  
 Db 1474 ---GNKLVIITDKAQSHEHETQSTFEKGQVVLQAGNDANILGNSVIDNGTQIQAGNHVR 1530  
 QY 1588 DSATTKMLAALKELDDAKPAEQHLIQHFSKADVVGDERYEAVRNKLVKLVROQAADSH 1647  
 Db 1531 IGTI-----QTSQSETHYQTKSGLSAGIG-----FTI---GSKTN 1565  
 QY 1648 SMLGSAHSTYNNLRINDGIVELLKHFDALPASSAKRLGEMW-----NNDPALK 1702  
 Db 1566 TQENQSQSNEHTGSTVSLGDDTTI-VAGKHYE-----QIGSTVSSPEGNNTIYAQ 1615  
 QY 1703 DIKQLQSTPSSASVSMELKQGLREQTEKAILDGKVGREVG-----LFQDNNLRV 1756  
 Db 1616 SIDIQAHHNKLNSNTTQTYEQKGLTVAFSSPVTD--LAQQAIAVAQSKQVQSKND-RV 1672  
 QY 1757 KVSVSQS--VSKSEGFNTALLGTSNSAAMSMERNITGTFNKYQDQNTPRFT-LEG 1813  
 Db 1673 NAMAANAGWQAYTGKSAQNLANGTTNAKQVS-----ISITYGEQN--RQTQVQA 1723  
 QY 1814 GIAQNPVQASALDLDKKEGLEMK 1838  
 Db 1724 NQAQSQIQAGGKTTLIATGAAEQS 1748

RESULT 14

ABU06028 standard; protein; 2514 AA.

AC ABU06028;

DT 23-OCT-2003 (revised)

DT 22-JAN-2003 (first entry)

XX

DE N. meningitidis vaccine antigen #16.

XX Vaccine; antigen; meningococcal disease; pathogenic bacteria; meningitis.

XX Neisseria meningitidis serogroup B.

XX WO200277648-A2.

XX 03-OCT-2002.

XX 22-MAR-2002; 2002MO-GB001399.

XX 22-MAR-2001; 2001GB-00007219.

XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.  
 FA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Robinson A, Gorringe AR, Hudson MJ, Bracegirdle P, West DM;

Oliver KJ, Kroll JS, Langford PR;

WPI: 2003-018958/01.  
 N-PSDB; ABX09865.

Identifying an antigen for manufacturing a vaccine against meningococcal infection, comprises contacting antibodies with polypeptides, detecting polypeptide-antibody complexes, and identifying bound polypeptides as antigens.

Claim 37; Page 156-164; 310pp; English.

The invention relates to identifying an antigen comprising: (a) obtaining antibodies against a commensal bacteria, or an extract from a commensal bacteria; (b) contacting the antibodies with polypeptides obtained from an expression library of either a commensal or a pathogenic bacteria; (c) determining whether the polypeptides bind to antibodies; and (d) (where a polypeptide binds to an antibody) identifying that polypeptide as an antigen. Also included are: (1) a method of preparing a vaccine composition, comprising identifying an antigen with the above method, and combining the antigen with a carrier; (2) a vaccine composition obtained by the above methods; (3) an antigen identified by the above methods; (4) a polypeptide encoded by all or a part of a nucleic acid sequence comprising the *Neisseria lactamica* DNA sequences detailed in the specification; (5) a vector comprising the nucleic acid molecule; (6) a method of preparing a composition for vaccination against infection by pathogenic bacteria, comprising: (a) obtaining a first antigen from a commensal *Neisseria*; (b) comparing the amino acid sequence of the first antigen with the amino acid sequence of the second antigen from a pathogenic bacteria, or comparing the sequence of a nucleic acid which codes for the first antigen with the sequence of the nucleic acid that codes for the second antigen; and if the first antigen is homologous to the second antigen or if the nucleic acid sequence for the first antigen is homologous to that of the second antigen, and (c) preparing a composition for vaccination against bacterial infection comprising the first antigen; and (7) an antibody that binds to the polypeptide antigen. The method is useful in screening commensal and pathogenic bacteria for previously unidentified vaccine antigens by identifying polypeptide antigens that bind to sera raised against commensal bacterial proteins. The polypeptide is useful as a vaccine antigen which may be used in the manufacture of a medicament for vaccination against meningococcal infection (e.g. meningitis). The present sequence represents an antigenic protein from the pathogenic bacteria *N. meningitidis*. (Updated on 23-OCT-2003 to standardise OS field)

Sequence 2514 AA;

Query Match 2.9%; Score 275; DB 6; Length 2514;

Best Local Similarity 18.8%; Pred. No. 1.5e-07;

Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY 7 GTEH-KAAVHTAAHPV-----GHGVALQCGSS-----SSPQNAAS 43  
 Db 37 GSAHVKSVPFGTTHAPVCRSNIFSLGFSCLAVGTANAFADGIADKAAPTQOAT 96  
 QY 44 LAEGKNRGKMPRIHQPSSTAADGISAAHQKSPSLRGCLGKTKFSRSPAGQPG----- 98  
 Db 97 ILQTGNG---IPQVNIOTPTTSAGVSVNQYAFDVGNEGAILNN--SRSTNTQTLGGWITQ 151  
 QY 99 -----TTHSK-----GATRLDRLARDGCGTQHEAAPPDAALTR 132  
 Db 152 NPMLARGEARVVVNQINSSSHSSQMNQYIEVGGRAEVVIANPAGIANGGGINASRTL 211  
 QY 133 SGGVKRENMDMAGRPVKG-----GSGED-----KVPTQQRHQLNNFGOMRQTMLS 180  
 Db 212 TTGQPYQAGDLGSGFKIRQGNVVIAGHGLDARDTDFTRILSYHSKIDAPVWQDVRVVG 271  
 QY 181 KMAHPASAGDRQLQHSPPHIPGSHHEIKKEPVGSTSKATTATAHADRVEIAQEDDDSEFOQ 240  
 Db 272 QNDVVAITGNA-----HSP-----ILNNAANTSNNTANNNGTHIFLEAIDTG----- 312  
 QY 241 LHOQRLARENPPEPPKPLGVATPISARFQPKLTAVAESVLEGTTTQSPKPSQMLKGS 300

Db 313 -----KLG-----GMVANKITLSTA-----EQAGIRNQQLFAS 342  
QY 301 GAGVTEPLAVTDKGLQLAPDNPALNTLLKQTLGKDTQHYLAHAS--SDGSHLLLDN 358  
Db 343 SGH-----VAIDANGRLVNSGTAAAN-----AKDNTAEKHVNRISQG-----VEN 385  
QY 359 KG-----HLFDIKSTATSYS-----VLHNSHFEIEIKGLAQAGTGSVSDGSGKI 404  
Db 386 SGTAVSQOQTQIHSQSIGTNTGTLSSGEILLHNS-----GSLKNETSGTI-----EAARL 435  
QY 405 SLGSGTQSHNKTMISQGEAHRSLTGIWOHPAGARPOGESIRLHDDKHLHILPELGW 464  
Db 436 AIDTDT--LNNQGLSOTG-----SOKLHI-----458  
QY 465 QSADKOTHSLSQADQCKYALKDNRTLQNLSD-----NKSSEKLVDKIKSYSDVQRGQ 518  
Db 459 -----DAQGM-----DNRGEMGLQDTAPTASGSSNOTGN-----SYNASFHS 498  
QY 519 VAILTDPGRHKMSIMPSLDASPEHISLSLHPADAHQGLLHKGSELEAQSVAISHGRV 578  
Db 499 TTTPTTATGTGTATVGSINAPT-----FAD-----GTIRTHGALDNGSGTIIANGQTD 547  
QY 579 VADSEKLSAALPKQDCGNELKWKAMPOHALDEHPCHDQIISGFFHDDHCOLNALVKN-- 637  
Db 548 VSAQCG--LNNAG--QIDIHQLNAG--SAFONHNG--TIIISDAVHIQAGSLNNQNGNI 598  
QY 638 NFRQCHACPLGNDHQFHPGWNLTALVID--NOLGLHHTNPEPEIILDMHGLGSLALQEGK 696  
Db 599 TTRFQ-----LEIETDQLDNAHGKLLSAEIALDLAVSGSLNNQNGE 638  
QY 697 LHYFDQLTKGWTGAESDCKOLKGLDGAAYLLKDGVEKRLINQSTSSIKHGHTENFSLP 756  
Db 639 IATNQUL-----IIHDGQSTAVIDNTWNTGTIQSGRDVAIQAK 675  
QY 757 HVNKEPEPDALQGLNK-----DDKAQAMAVIGVNVKYLALTEKGDIRSFIQKPGTQOLE 810  
Db 676 SLSN-----NGTLAANDKLDIALQDDFYVERNIVAGNE--JLSLSTRGKLSHNTTQAGKRIR 730  
QY 811 RPAQTLISREG-----ISGELKDIIHVDHKQNLIALYTHEGEVPHQPREAWQNGAESSSHKL 866  
Db 731 IKANNDNAAGNIQSGTTDICTQHN-----LTNRGLIDGQ-----767  
QY 867 LPOSEKLSKLDMSH-----EHPKIATFEGSQHQLKAGGWHAYAAPERGPLAVGTS 918  
Db 768 -----QTKIOAGOMNIGTGTIRYDNTAIAATRLDNDENGTA--AIAARENILGTV-- 819  
QY 919 GSQTVENRLMQGVKVPISGLTVKLSAQTGCTGAEGKVKSSKTSERIRAYAFNPTMS 978  
Db 820 -----QLNRENLSIYSGNDMAVGALDNTNQATGKAOR-----853  
QY 979 TPRIKNAAYATQHGQREGKPLXYEMOGALIKOL-----DAHNVHRNAP 1024  
Db 854 -----IHNAGATIEAAGKMRGLGVEKLHNTNEHLATQLVETGREHIVDYEAFGREHLLREGT 909  
QY 1025 QPDL-----QSKLETLLDGEH-----GAELNDMK 1049  
Db 910 QHELGSVYNDESDDLTPDGAHENWHKYDEKVTQKTQVQTAPAKIISGNDLTDIGK 969  
QY 1050 R--FRDELEQSATRSVTVLGHQHG-----VLKSNGEINSEFKPS--PGKALYQSFN 1096  
Db 970 EVENTDSQIIAGNLIVQTEKDLHNEQTFGEKKVPSENGKLSHYREKHGR-----DSTG 1026  
QY 1097 VNRSGLDKSLQQAHAATPPSAESKLSQMLGHFVSAGVDMSHQKGEIPLGRORDPNDKT 1156  
Db 1027 HSEQNTLPPEITRNISLGSFAVESHRKALSHAPSQGTLPQSNQ---ISLPTVNSFT 1083  
QY 1157 ALTKSRL--ILDVTYIGELHELADAK-----LYSDHKPDA---DQIKQLRQPDFTREXR 1207  
Db 1084 PLPSSLYIINPVNKGVLVETDPRFANYROWLGSYMLDSLKLDPNLHKLRLGDGYYEQR 1143  
QY 1208 YENPNKYVYDMGFTHNKALEANYDAVKAFINAFKKEHGHVNTTTRTVLESQGSAAELAKK 1267  
Db 1144 LINEQIAELT--GHRRLDGYQNDDEQFKALMDNGATAARSNNLSVGIAL-----SARQVQ 1197

QY 1268 LKNTLLSLDSGESMSPRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSFRTS 1327  
Db 1198 LTSDIVMLVQKE-----VKLPDGGTQIVLPQVVRVKNKGIDGKGAL-----LSGSNTQ 1247  
QY 1328 GGLNVSGRGGGVSGNIMVATGHDVMPYMTGK---KTSA-----GNASDWLSAKHKIS 1377  
Db 1248 INVSGSLKNSGTIAGRNALITNTDLNIGGRIHAQKSAVTATQDINNIGMLSAEQTL 1307  
QY 1378 PDLRIGAAV--SGTLQCTLQNSLKFKLTEDELPG--FIHGLTHGTLT-----PAELLQKIE 1430  
Db 1308 ---LWAGNNINSQTTASSQNTQSSSYLDRMAGIYITGKEKGLAAQAGKDINIAGQIS 1365  
QY 1431 HQMKQSKLTFSDVTSANLTL-----RAGINLINEDGSKPNGVTARYSAGLSASA----- 1479  
Db 1366 NQSEQG--QTRLOAGRDINDLTQVTSKHQATHPADNHHVIRGSTNEVGSSITQKGDVTL 1424  
QY 1480 --NLAAGSRSTTSQFGSTTSASNRPFLNGVAGANLTAALGVVAHSSTHEKPVGI 1337  
Db 1425 GNNLNAAAABVSSANGTL--AVSAKND-----IN--ISAGINTT--HVDDASKHTGRSGG-- 1473  
QY 1538 FPAPTSTNVSAALALDNRTSQSISLELKRAEPVTSNDISELTSTL-----GKHFK 1587  
Db 1474 ---GNKLVTIDKAAQSHHETASQSTFEGKQVVLQAGDANILGSNVISDNGTQIQAGNHVR 1530  
QY 1588 DSATTTMLAALKELDDAKPAEQHLILQOHFSAKDVVGDERYEAVRNLLKLVIRQQAADSH 1647  
Db 1531 IGTI-----QTSQSETVHOTQKSLMSAGIG-----FTI---GSKTN 1565  
QY 1648 SMELGSASHSTTVNNLSRINDGIVELLHKKHFDPAALPASSAKLGEWM-----NNDPALK 1702  
Db 1566 TQENQSQSNEHTGSTVGS�KGDTTI--VAGKHYE-----QIGSTVSSPEGNTTIYAQ 1615  
QY 1703 DIILQLOSTPFSSASVSMELKQGLREQTEKAILDKGVGREVG-----LFODRNLRV 1756  
Db 1616 SIDIQAHNKLNSNTTQVEKGLTVAFSPVTD--LAQAIIVAOSSKQVQSKND--RV 1672  
QY 1757 KSVVSQS--VSKSEGTNPALLGTSNSAAMSERNICTINPKYQDQNTPRRFT--LEG 1813  
Db 1673 NAMAANAAGWQAYQTKSAQNLANGTTNAQVS-----ISITYGEQQN--RQTTQVQA 1723  
QY 1814 GTAQANPQVASALTDLKKEGLENKS 1838  
Db 1724 NQAQASQIQAGGKTTLIATGAAEQS 1748

## RESULT 15

ABU38197  
ID ABU38197 standard; protein; 2514 AA.

XX AC ABU38197;

XX AC AC

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #23724.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Neisseria meningitidis.

XX PN WC200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:15:47 ; Search time 35.3283 Seconds  
(without alignments)  
2685.909 Million cell updates/sec

Title: US-09-596-784-2  
Perfect score: 9448  
Sequence: 1 MELKSLGTEHKAHVHTAAH.....NPQVASALTDLKKEGLEMK 1838

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/ECTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9448	100.0	1838	3	US-09-120-663-2
2	9448	100.0	1838	4	US-09-431-614-8
3	1041	11.0	201	4	US-09-350-852A-1
4	257.5	2.7	10182	4	US-09-134-001C-3159
5	238.5	2.5	2504	4	US-09-328-352-5821
6	238	2.5	1589	4	US-09-543-681A-4998
7	236	2.5	3241	4	US-09-841-786-1
8	233	2.5	3696	4	US-09-134-001C-5080
9	216	2.3	2048	4	US-09-268-347-48
10	211.5	2.2	2680	4	US-09-489-039A-7973
11	209.5	2.2	1912	1	US-08-409-995-4
12	209.5	2.2	1912	3	US-08-685-467-4
13	207.5	2.2	2142	4	US-09-540-236-3459
14	205.5	2.2	2123	3	US-08-968-685A-10
15	204.5	2.2	1992	4	US-08-621-944A-3
16	204.5	2.2	1992	4	US-08-945-567D-3
17	203.5	2.2	1780	1	US-08-769-309A-5
18	203.5	2.2	1780	3	US-08-994-570-5
19	201	2.1	2137	4	US-09-134-001C-4463
20	201	2.1	2353	3	US-08-377-155-33
21	201	2.1	2353	3	US-08-913-942-4
22	201	2.1	2353	4	US-09-669-974-33
23	201	2.1	2353	4	US-09-797-862-33
24	197	2.1	2291	4	US-09-252-991A-21854
25	196	2.1	2354	4	US-09-268-347-47
26	195	2.1	2042	4	US-09-077-098A-6
27	195	2.1	2568	4	US-09-866-108A-3

28 194 2.1 2954 4 US-09-150-867-1 Sequence 1, Appli  
29 193.5 2.0 2314 4 US-09-268-347-49 Sequence 49, Appli  
30 190.5 2.0 1536 1 US-08-038-682-2 Sequence 2, Appli  
31 190.5 2.0 1536 1 US-08-302-832-2 Sequence 2, Appli  
32 190.5 2.0 1536 2 US-08-530-198-2 Sequence 2, Appli  
33 190.5 2.0 1536 2 US-08-469-880-2 Sequence 2, Appli  
34 190.5 2.0 1536 2 US-08-728-470-2 Sequence 2, Appli  
35 190.5 2.0 1536 2 US-08-617-697-2 Sequence 2, Appli  
36 190.5 2.0 1536 3 US-08-719-641-2 Sequence 2, Appli  
37 190 2.0 1315 4 US-09-252-991A-22746 Sequence 22746, A  
38 190 2.0 2411 4 US-09-268-347-36 Sequence 36, Appli  
39 188.5 2.0 1420 4 US-09-125-635-4 Sequence 4, Appli  
40 188.5 2.0 1536 4 US-09-206-942-67 Sequence 67, Appli  
41 188.5 2.0 2039 4 US-09-077-098A-7 Sequence 7, Appli  
42 187 2.0 1600 2 US-08-617-697-10 Sequence 10, Appli  
43 186.5 2.0 2482 1 US-08-328-254-6 Sequence 6, Appli  
44 185.5 2.0 1761 4 US-09-489-039A-11234 Sequence 11234, A  
45 184.5 2.0 1833 4 US-08-621-944A-4 Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-09-120-663-2  
; Sequence 2, Application US/09120663  
; Patent No. 6228644  
; GENERAL INFORMATION:  
; APPLICANT: Bogdanove, Adam J.  
; APPLICANT: Kim, Jihyun Francis  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Beer, Steven V.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM  
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/120,663  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,105  
; FILING DATE: 06-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1661  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1600  
; TELEFAX: (716) 263-1304  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1838 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-120-663-2

Query Match 100.0%; Score 9448; DB 3; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MELKSLGTEHKAHVHTAAHNPVGHGVALOOGSSSSPQNAASAAIAAEGKNGKMPRIHOP 60  
 1 MELKSLGTEHKAHVHTAAHNPVGHGVALOOGSSSSPQNAASAAIAAEGKNGKMPRIHOP 60  
 61 STAADGISAHQKQKSFSLRGCLGTGKFFRSAPQGGPQGTTHSGATLRLDILLARDGGETQH 120  
 61 STAADGISAHQKQKSFSLRGCLGTGKFFRSAPQGGPQGTTHSGATLRLDILLARDGGETQH 120  
 121 EAAAPDAARLTRGGVKRRNMDMAGBPMVKGGSGEDKVPQOKRHLNNFGQMRQTMLS 180  
 121 EAAAPDAARLTRGGVKRRNMDMAGBPMVKGGSGEDKVPQOKRHLNNFGQMRQTMLS 180  
 181 KMAHPASANAGDRLQSHPPHICSHHHEIKEEPVGSSTSKATTAHADRVETIAQEDDDSEFQQ 240  
 181 KMAHPASANAGDRLQSHPPHICSHHHEIKEEPVGSSTSKATTAHADRVETIAQEDDDSEFQQ 240  
 241 LHQORLARERENPPQPKLGVAATPISARFQPKLTAVAESVLEGGTDTTQSPKQSMKLGKS 300  
 241 LHQORLARERENPPQPKLGVAATPISARFQPKLTAVAESVLEGGTDTTQSPKQSMKLGKS 300  
 301 GAGVTPAVTLDKGLQAPDNPALNTLLKQTKDQTOHYLAHASSDGSQHLLLDNKG 360  
 301 GAGVTPAVTLDKGLQAPDNPALNTLLKQTKDQTOHYLAHASSDGSQHLLLDNKG 360  
 361 HLPDIKSTATSYVLHNSHPGEIKGLAQAGTGSVSDGSGKISLGSQTQSHNKTMLSQ 420  
 361 HLPDIKSTATSYVLHNSHPGEIKGLAQAGTGSVSDGSGKISLGSQTQSHNKTMLSQ 420  
 421 PGSAHRSLLTGIWQHPAGARPOGESIRLHDDKIHLHPELVQWQADKQTHSOLSRQAD 480  
 421 PGSAHRSLLTGIWQHPAGARPOGESIRLHDDKIHLHPELVQWQADKQTHSOLSRQAD 480  
 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYVDVQGOVALTDTTPGRHKMSIMPSLDAS 540  
 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYVDVQGOVALTDTTPGRHKMSIMPSLDAS 540  
 541 PEGHISLSLHFAHQAQGLLHGKSELEAQSVASHGRVLVADSEGLFSAAPKQGDGNE 600  
 541 PEGHISLSLHFAHQAQGLLHGKSELEAQSVASHGRVLVADSEGLFSAAPKQGDGNE 600  
 601 KMKAMPOHALDEHFGHGHQISGFFHDDHQLNALVKNFRQOHAACPLGNHOFHPGNWLT 660  
 601 KMKAMPOHALDEHFGHGHQISGFFHDDHQLNALVKNFRQOHAACPLGNHOFHPGNWLT 660  
 661 DALVIDNOLGLHHTNPRPHETLDMGHLGSLAQSGKLHYPDQTKGTGABSDCKQLKKG 720  
 661 DALVIDNOLGLHHTNPRPHETLDMGHLGSLAQSGKLHYPDQTKGTGABSDCKQLKKG 720  
 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 721 LDGAAYLLKDGVEKRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
 781 AVIGVKNYALATEKGDIRSFOIKPQTOOLERPQATLSREGISGEIKOIHVDHKNLYALT 840  
 781 AVIGVKNYALATEKGDIRSFOIKPQTOOLERPQATLSREGISGEIKOIHVDHKNLYALT 840  
 841 HEGEVFHPQRAWONGAESSWHKLALPOSESKLSLMSHEHKPIATFFEDGSOHLKAG 900  
 841 HEGEVFHPQRAWONGAESSWHKLALPOSESKLSLMSHEHKPIATFFEDGSOHLKAG 900  
 901 GWHAYAAPERGLAVGTSGSTVFNRMLQGVKGKVIPEGSLTVKLSAQGTGMTGAERGV 960  
 901 GWHAYAAPERGLAVGTSGSTVFNRMLQGVKGKVIPEGSLTVKLSAQGTGMTGAERGV 960  
 961 SSKFSERIRAVAFNPTWSTPRPIKNAAYATOHGQWQREGIKPLXYEQGALIKQIDAHNVR 1020  
 961 SSKFSERIRAVAFNPTWSTPRPIKNAAYATOHGQWQREGIKPLXYEQGALIKQIDAHNVR 1020  
 1021 HNAPODQLQSKLETLDLGEHGAELLNDMKRFRDELEQSAATRSVTVLGQHQGVLSNGEIN 1080  
 1021 HNAPODQLQSKLETLDLGEHGAELLNDMKRFRDELEQSAATRSVTVLGQHQGVLSNGEIN 1080

1081 SEFKPSFGKALVQSFNVRNSGQDLKSLOQAVHATPPSAESKLOSMLGHPVFSAGVDMSHQ 1140  
 1081 SEFKPSFGKALVQSFNVRNSGQDLKSLOQAVHATPPSAESKLOSMLGHPVFSAGVDMSHQ 1140  
 1141 KGEIPLGRORDPNDKATLTKSRLILDTVTIGELHELADRAKLVSDHKPADQIKQLRQQF 1200  
 1141 KGEIPLGRORDPNDKATLTKSRLILDTVTIGELHELADRAKLVSDHKPADQIKQLRQQF 1200  
 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260  
 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260  
 1261 SAEKAKLKNLTLSDSGESMSPRSYGGVSVFVPTLSKKYVPVPIPGAGITLDRAYN 1320  
 1261 SAEKAKLKNLTLSDSGESMSPRSYGGVSVFVPTLSKKYVPVPIPGAGITLDRAYN 1320  
 1321 LPSFRTSGGLNVSGFRDGGVGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380  
 1321 LPSFRTSGGLNVSGFRDGGVGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380  
 1381 RIGAAVSGTLQGTQLNSLKFPLTEDELPFGLTHGLTHGLTPAEILLQKGIHOMKQGSKL 1440  
 1381 RIGAAVSGTLQGTQLNSLKFPLTEDELPFGLTHGLTHGLTPAEILLQKGIHOMKQGSKL 1440  
 1441 FSVDTSANLDRAGINLINEDGSKPENGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500  
 1441 FSVDTSANLDRAGINLINEDGSKPENGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500  
 1501 ASNNRPTFLNGVAGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
 1501 ASNNRPTFLNGVAGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
 1561 SLEKRAEPTVSTNDISELTSTLKGFKDSATTWMLAALKELDDAKPAEQHLHQHFSAK 1620  
 1561 SLEKRAEPTVSTNDISELTSTLKGFKDSATTWMLAALKELDDAKPAEQHLHQHFSAK 1620  
 1621 DVVGDERYEAVRNLKLVIRQOARADSHMELGASHSTTYNNLSRINNDGIVELLKHFD 1680  
 1621 DVVGDERYEAVRNLKLVIRQOARADSHMELGASHSTTYNNLSRINNDGIVELLKHFD 1680  
 1681 AAPASSAKELGEMMNDPALKDIIKQLOSTPSSASVSMELKDGLREOTEKAILDGKVG 1740  
 1681 AAPASSAKELGEMMNDPALKDIIKQLOSTPSSASVSMELKDGLREOTEKAILDGKVG 1740  
 1741 REEVGLVFQDRNNLRVKSVSQSVKSEGFPNTALLGTSNSAAMSVERNIGTINPKY 1800  
 1741 REEVGLVFQDRNNLRVKSVSQSVKSEGFPNTALLGTSNSAAMSVERNIGTINPKY 1800  
 1801 ODQNTPRFRFTLEGGIAQANPQASALTDLKKEGLEMS 1838  
 1801 ODQNTPRFRFTLEGGIAQANPQASALTDLKKEGLEMS 1838

RESULT 2  
 US-09-431-614-8  
 ; Sequence 8, Application US/09431614  
 ; Patent No. 6624139  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wei, Zhong-Min  
 ; APPLICANT: Schading, Richard L.  
 ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS  
 ; TITLE OF INVENTION: RESISTANCE  
 ; FILE REFERENCE: 21829/41 (EBC-003)  
 ; CURRENT APPLICATION NUMBER: US/09/431,614  
 ; CURRENT FILING DATE: 1999-11-02  
 ; EARLIER APPLICATION NUMBER: 60/107,243  
 ; EARLIER FILING DATE: 1998-11-05  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: Patent in ver. 2.0  
 ; SEQ ID NO 8  
 ; LENGTH: 1838  
 ; TYPE: PRT  
 ; ORGANISM: Erwinia amylovora

US-09-431-614-8

Query Match 100.0%; Score 9448; DB 4; Length 1838;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MELKSLGTEHKAHVTAHNPVGHVALQOGSSSSPQNAASLAABKQNRKMPRIHQ	60
Db	1	MELKSLGTEHKAHVTAHNPVGHVALQOGSSSSPQNAASLAABKQNRKMPRIHQ	60
Qy	61	STAADGISAHQKQKSFSLGCLGTFKFSRPAQOGPTTHSKGATLRLDARDGETQH	120
Db	61	STAADGISAHQKQKSFSLGCLGTFKFSRPAQOGPTTHSKGATLRLDARDGETQH	120
Qy	121	EAAPDAARLRSRGVGRNMDMAGRPMVKGSGEDKVPYQQRHQLNFMQFQMTLS	180
Db	121	EAAPDAARLRSRGVGRNMDMAGRPMVKGSGEDKVPYQQRHQLNFMQFQMTLS	180
Qy	181	KWHPASANAGDRLOHSPPHIPGSHHEIKKEPVGSGTAKTAAHADRVEIAEDDDSEFQ	240
Db	181	KWHPASANAGDRLOHSPPHIPGSHHEIKKEPVGSGTAKTAAHADRVEIAEDDDSEFQ	240
Qy	241	LHQRLAREENPPQPKLGAVATPISARFQPKLTAVAESVLEGTDTTOSPLKPSQMLKGS	300
Db	241	LHQRLAREENPPQPKLGAVATPISARFQPKLTAVAESVLEGTDTTOSPLKPSQMLKGS	300
Qy	301	GAGVTPLAUTLDKGLQAPNPPALNTLLKQTLGKDTQHYLAHHAASDGSQHLLDNKG	360
Db	301	GAGVTPLAUTLDKGLQAPNPPALNTLLKQTLGKDTQHYLAHHAASDGSQHLLDNKG	360
Qy	361	HLFDIKSTATSYSVLHNSHPHIEIKGLAQAGTGSVVDGSKGISLGSQTSHNKTMLSQ	420
Db	361	HLFDIKSTATSYSVLHNSHPHIEIKGLAQAGTGSVVDGSKGISLGSQTSHNKTMLSQ	420
Qy	421	PGEAHRSLITGIWOHPAGAARPOGESIRLHDDKTHILHPELGVMSADKDTSHQSRD	480
Db	421	PGEAHRSLITGIWOHPAGAARPOGESIRLHDDKTHILHPELGVMSADKDTSHQSRD	480
Qy	481	GKLYALKDNRTLQNLSDNKSSEKLVDKISYVDRQGVAILTDTGRRHMSIMPSLDAS	540
Db	481	GKLYALKDNRTLQNLSDNKSSEKLVDKISYVDRQGVAILTDTGRRHMSIMPSLDAS	540
Qy	541	PESHISILHPADAHQGLHKGSELEAQSVAISHGRLVVADSEKGLFSAALPKQGDGNE	600
Db	541	PESHISILHPADAHQGLHKGSELEAQSVAISHGRLVVADSEKGLFSAALPKQGDGNE	600
Qy	601	KWKAMPOHALDEHFGHDHIOISGFPHDDHQLNALVKNNFRQOHCACPLGNDHOFHPGNLT	660
Db	601	KWKAMPOHALDEHFGHDHIOISGFPHDDHQLNALVKNNFRQOHCACPLGNDHOFHPGNLT	660
Qy	661	DALVIDNQLGLHHTNPEPHEILDGMHLGSLALQEGKLYHFDQLTKGWTGAESDCKQLKKG	720
Db	661	DALVIDNQLGLHHTNPEPHEILDGMHLGSLALQEGKLYHFDQLTKGWTGAESDCKQLKKG	720
Qy	721	LDGAAYLLKQGEVRLINQSTSSIKGTENVFLPHVRNKPBPQDALQGLNKDDKAQAM	780
Db	721	LDGAAYLLKQGEVRLINQSTSSIKGTENVFLPHVRNKPBPQDALQGLNKDDKAQAM	780
Qy	781	AVIGWNKYLALTEKGDIRSFOIKPTQOLRPAOTLSREGISGELKDTHVDHKNLYALT	840
Db	781	AVIGWNKYLALTEKGDIRSFOIKPTQOLRPAOTLSREGISGELKDTHVDHKNLYALT	840
Qy	841	HEGEVFPHPREAWNGAESSSWHKLAIPQSESKLSLSDMSHEHKPIATFEDGSHQKAG	900
Db	841	HEGEVFPHPREAWNGAESSSWHKLAIPQSESKLSLSDMSHEHKPIATFEDGSHQKAG	900
Qy	901	GWHAYAPERGLAVGTSGSTVFNRLMQGVKQVIPSGLTVKLSAQGTGMTCAEGRKV	960
Db	901	GWHAYAPERGLAVGTSGSTVFNRLMQGVKQVIPSGLTVKLSAQGTGMTCAEGRKV	960
Qy	961	SSKFSERIRAVAFNPTMSTPRPIKNAAYATQHGQREGLKPLIYEMQGALIKQDAHNR	1020
Db	961	SSKFSERIRAVAFNPTMSTPRPIKNAAYATQHGQREGLKPLIYEMQGALIKQDAHNR	1020

Qy	1021	HNAPODLQSKLETILDGELGAEILNDMKRFRDELEQSATRSVTVLGQHGVKNSGEIN	1080
Db	1021	HNAPODLQSKLETILDGELGAEILNDMKRFRDELEQSATRSVTVLGQHGVKNSGEIN	1080
Qy	1081	SEFKPSPKALVQSFNNRSGDLSKSLQQAQVHATPPSAESKLQSMGLHFVSAGVDMSHQ	1140
Db	1081	SEFKPSPKALVQSFNNRSGDLSKSLQQAQVHATPPSAESKLQSMGLHFVSAGVDMSHQ	1140
Qy	1141	KGEIPLGRORPNDKALTAKSLILDTVTIGELHELADKAKLVSDHKPDADQIKOLRQOF	1200
Db	1141	KGEIPLGRORPNDKALTAKSLILDTVTIGELHELADKAKLVSDHKPDADQIKOLRQOF	1200
Qy	1201	DTLREKRYESNPVHYITDMGFTTHNKALEANTDAVKAFINAFKKEHGVNLTTRTVLESOG	1260
Db	1201	DTLREKRYESNPVHYITDMGFTTHNKALEANTDAVKAFINAFKKEHGVNLTTRTVLESOG	1260
Qy	1261	SAELAKKUNTLILLSGSEMSFSRSGYGGVSTVFPVTLSSKKVPVPIPGAGITLDRAYN	1320
Db	1261	SAELAKKUNTLILLSGSEMSFSRSGYGGVSTVFPVTLSSKKVPVPIPGAGITLDRAYN	1320
Qy	1321	LSFRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD	1380
Db	1321	LSFRTSGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD	1380
Qy	1381	RIGAAVSGTLQGTQLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIHQHOMKQSKIT	1440
Db	1381	RIGAAVSGTLQGTQLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIHQHOMKQSKIT	1440
Qy	1441	FSVDTSANLDIRAGINLNEGSKPVGTVARVAGLSASANLAAGSRERSTTSGQFGSTTS	1500
Db	1441	FSVDTSANLDIRAGINLNEGSKPVGTVARVAGLSASANLAAGSRERSTTSGQFGSTTS	1500
Qy	1501	ASNNRPFLNCGVAGANLTAALGVASHSTHEGKPVGIPPAFTSTNVSAALDNRTSOSI	1560
Db	1501	ASNNRPFLNCGVAGANLTAALGVASHSTHEGKPVGIPPAFTSTNVSAALDNRTSOSI	1560
Qy	1561	SLELKRAEPTVTSNDISELTSTLGHFKDSDATTKMLAALKELDDAKPAQLHLOOHFSAK	1620
Db	1561	SLELKRAEPTVTSNDISELTSTLGHFKDSDATTKMLAALKELDDAKPAQLHLOOHFSAK	1620
Qy	1621	DVGDREYEAURNKLVIROQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHGHFD	1680
Db	1621	DVGDREYEAURNKLVIROQAADSHSMELGSASHSTTYNNLSRINNDGIVELLHGHFD	1680
Qy	1681	AALPASAKLGENMNDPALKDIKQIQLQSTPPSSASVSMELKGLRQTEKAILDGKVG	1740
Db	1681	AALPASAKLGENMNDPALKDIKQIQLQSTPPSSASVSMELKGLRQTEKAILDGKVG	1740
Qy	1741	REEVGVLFDORNLRVKSVSQSVKSEGFNTALLGTSNSAAMSWERNIGTINPKYG	1800
Db	1741	REEVGVLFDORNLRVKSVSQSVKSEGFNTALLGTSNSAAMSWERNIGTINPKYG	1800
Qy	1801	QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMSK 1838	
Db	1801	QDQNTPRRFTLEGGIAQANPQVASALTDLKKEGLEMSK 1838	

RESULT 3

US-09-350-852A-1  
; Sequence 1, Application US/09350852A  
; Patent No. 6596509  
; GENERAL INFORMATION:  
; APPLICANT: Bauer, David W.  
; APPLICANT: Beer, Steven V.  
; APPLICANT: Bogdanove, Adam J.  
; APPLICANT: Collmer, Alan  
; APPLICANT: Ham, Jong Hyun  
; TITLE OF INVENTION: RECOMBINANT CONSTRUCTS AND SYSTEMS FOR SECRETION OF  
; FILE REFERENCE: 19603/2061  
; CURRENT APPLICATION NUMBER: US/09/350, 852A  
; CURRENT FILING DATE: 1999-07-09

```
; PRIOR APPLICATION NUMBER: 60/092,357
; PRIOR FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 201
; TYPE: PRN
; ORGANISM: Erwinia amylovora
US-09-350-852A-1

Query Match 11.0%; Score 1041; DB 4; Length 201;
Best Local Similarity 100.0%; Pred. No. 7.2e-66;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSPQNAASAAAEKGRGKMPRIHQ 60
Db 1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSPQNAASAAAEKGRGKMPRIHQ 60

QY 61 STAADGISAHQOKKFSRGLGCTKKFSRSPQGGPGTTHSKGATLRDLLARDGGETQ 120
Db 61 STAADGISAHQOKKFSRGLGCTKKFSRSPQGGPGTTHSKGATLRDLLARDGGETQ 120

QY 121 EAAAPDAALTRSGGVKRRNMDMAGRPVKGSGEDKVPQOKRHQNLNPFQGMQRTMLS 180
Db 121 EAAAPDAALTRSGGVKRRNMDMAGRPVKGSGEDKVPQOKRHQNLNPFQGMQRTMLS 180

QY 181 KMAHPASANAGDRLOHSPPHI 201
Db 181 KMAHPASANAGDRLOHSPPHI 201

RESULT 4
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match 2.7%; Score 257.5; DB 4; Length 10182;
Best Local Similarity 17.7%; Pred. No. 1.1e-07;
Matches 396; Conservative 319; Mismatches 805; Indels 723; Gaps 90;

QY 33 SSSSPQNAASL-AEGRKGRKMPRIHQPTAADGISAHQOKKFSRGLGCTKKFSRS 91
Db 6262 ATQSVQNAEQALHGAELNQDK-----QTSSTELDGLTDLTDAORE-KLREQINTNSRDD 6316

QY 92 APQ--GPGTTHSKGATLRDLLARDG-----ETOHEAAPDAAR----- 129
Db 6317 IKOXIEQKALNDAMKLEQVAKQDGVHANSDYTNESQAQDAVNNALKAQEDINNSS 6376

QY 130 -----LTRSGGVKRRNMDMAGRPVKGSGEDKVPQOKRHQNLNPFQGMQRTMLS 181
Db 6377 NPENLAQDITNALNNIKQAQNLHGAOKLIQ-----QDKNTTNQAIGNLNLNPFQOKDALIQ 6432

QY 182 MAHPASA-----NAGDRLOHSPPHIPGSHHEIK----- 209
Db 6433 AINGATSRDQVAEKLEAEALDEAMKQLEDQVQDDQISNSPFINESDKQTYNDKIQ 6492
```

```
QY 210 --EBPVGSTSKAT-----TAHADRVELAQEDDSEFQOLHQOBLARE 249
Db 6493 AAKEIINQTSNPTLKQKIADTLQNIKDAVNNLHGQD-KLAQSKQDANNQNLHDLDTTEE 6551

QY 250 RENPPQPKLGVAATPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKSGAGVTPPLAV 309
Db 6552 QKN-----HFKP-----LINNADTRDEVNK----- 6571

QY 310 TLDKGKQLAPONPPALNTLLKQTLGKDTQHYLAHSSDGSQHLLLDNKGHLFDTKSTA 369
Db 6572 -----OLEIAKQLNGDMSTLHKVINDKDOIQLHLSNVINADNDKKQYDN-----AIK 6618

QY 370 TSYSVLHNHPGHIKLAQAGTGSVSDGSKSISLGSTQSHNKTMLSQPGEAHRSL 429
Db 6619 EAEIDLHN-HPDITLDHKAQ-----DLNKIDQAHNEL- 6650

QY 430 TGIWQHAPGAAR-----PQGESIR--LHDDKIHLHLPGLGWQS 466
Db 6651 -----NGESRFKQALDNALNDIDSLSLNVFORQTVKDNINHVITLESQAQEL----- 6698

QY 467 ADKDTSHSQRQADGKLYALKD-----NRTQLNSDNKSEKLVDKI----- 508
Db 6699 -----QKAKELNDAMKARDSIMNQEQIIRKNSNYTNEDLAQNAYNHADVKNHILIGE 6751

QY 509 -----KSYSDVQRGOVALTDTTPGRHKMSIMPSLD 538
Db 6752 DNATMDPQIIOATODINTAINGLNGDQKIQDAKTDAKQITNFTGLTEPQKQAL----- 6806

QY 539 ASPESHISLSLHFAHAGHGLHGK---SELEASQVAISHGRLVAD-----SEGLKF 587
Db 6807 ---ENIINQOTSANVAKQLSHAKFLNGKMEELKVAVAKASLVRSNSYINEDVSEKEAY 6863

QY 588 SAAIPKQGDGNE-LKMKAMPQHALDEHFGHDHOISGFFHDDHGLQNLNVLKNNFRQQAACP 646
Db 6864 EQAIK---QGEIINSENNTISSTDINRTIOBNDAEQNLHGQ-----NKLRAQAEI- 6913

QY 647 LGNDHGFHPGWNUTDALIVNQILGLHHTPEPEHILDMGHLGSLALQEGKHLFYDQITKG 706
Db 6914 AKNEIQNLGNSAQITKLIQDITRTTKPAVTQKLEAKAIQNAQMOQLKQSIADK----- 6969

QY 707 WTGAESDCKLKKGLDGAAYLLKDGVEVKRLNIN-----QSTSSIKHGTENVFSLPHVN 760
Db 6970 -----DATLSNSNYNEDESEKKLAYDNVAVSOAQLINQNLNPTWIDISNIQAITQ 7018

QY 761 KP-EPGDALQGLNKDDKAQAMAVIGNVKYLALTEKGD-----IRSFQIKPGTQOLBERPA 813
Db 7019 KVIQAKDSLHGAANKLAQNAQDSNLIINQSTNLNDKQKQALNDLINHAQTKQQAEEIIAQ 7078

QY 814 QTLREGISGELKDIDVD-----HKQNLIALTHEGEVHQPRAWQNGAESSSW----- 862
Db 7079 NKLNNE--MGTLLKTLVEEQSNVHQOSKY--INEDPQVQNIYNDISIQKREILNGTTDDVLN 7135

QY 863 -HKALPQSESCLKSLDMSHEHKPIATFEDGSQH-----QLKAGGWHAAYAABERG 911
Db 7136 NKKIADAIQNLHTKNDLHGDKLQKAAQADATNELNVLTLNNSQROSEHDEINSASRT 7195

QY 912 PLAVGTSGSTQVFNRLMGQGVKVPISQGLTVKLSAQGTGGTGAEGRKVSKSFGERIRAY 971
Db 7196 EVSNDLNHAKAL-NEAMRQLENEVALENSVK-KLSDFIN-----EDEAAQAEVSNALQKA 7248

QY 972 --AENPTWST---PRPKNAAYATOHQWQREGIKPLIYEMQALIKOLD---AHN----- 1018
Db 7249 KDIINGVPSLTKATTEDALLEQNARESLHGQKLOEAKQNAQVAEIDNLQALNPGQVL 7308

QY 1019 -----VRHNAPOPDLQSKLETLDLGEHGAELINDMKFRDELEQASATRSVTVLQHQGV 1073
Db 7309 AEKTLVNAQSTKPEQVEAL-----QKAKELNEAMKALKTEINKK-----EQIK 7351

QY 1074 KSGNEINSEFKPSGKALVOSFN--VNRSGQ-----DLKSLQQAQVHATPPSAESKL 1123
Db 7352 ADSRYVNAD-----SGLQANYNSALNYGSIATTOPPELNKQVINRATQTIKTAENNL 7405

QY 1124 --QSMLGHPVSAG-VDSHQKGEIPLGFRQDPNDKALTLSRLILDTVTIGELHDLAKA 1180
```





797 IRSFQIKPGTQOLRPAQTLSREGISGELKDIHVHDKQNYALTHEGEVHFHPREAWQNG 856  
1393 ITKEGINAGSKQITNVADGINAKDAVNSQLDLAAKQ--ATDDAAVKYD-----1441  
857 ABSSWHKIALPOSE-SKLKSLDMSHEHPKPIAFEDGSO-HQLKAG-----GHWAAPE 909  
1442 -DAKTKDKVTLKGGDGTVDNVKAGHSISSTKEAVNGSQIHKISNSIKSIGGNTVVPND 1500  
910 RGLA---VGTSGSQTIVNELMQ-----GVKQKVIPIGSLTVKLSAQTGMTGAEGRKV 960  
1501 -GSLTNNIGGTGKNINDAISEVKNATKAKTIVTEGDNIVVVKETVNDKGSNTYE---V 1556  
961 SSFESERIRAYANPTM--STPRPIKAAVATQHGQWQREGKLPYEMQCALIKQLDAHN 1018  
1557 ATKKDLTNSVTTGDTFVNNGLTIKDGPSITKDGINA-SGKQITNVADG--INAKDAVN 1613  
1019 VRH-----NAPQPDLOSKEFTL--DLGEHGAEL--INDMKR-----1050  
1614 VDQLTVKENLNGRIITDNNQLNDAKDLGNQIADTNKNLNDAKDLGNQITDNTNTKINT 1673  
1051 PRDEQSATRSVTVLQHGQVLKSNSEKPKSPGKALVQSFN-VNRSQDLSKSLQ 1109  
1674 TKQLTQITQINDTTELNTTNTKT--ELNSKIDSTTELENKLNFGNSGNDVHRKLG 1731  
1110 QAVH-----ATPPSABSKLQSMGLHFVSVAGVDMSHQK---GEIPLGRQDPNDKALT 1159  
1732 EKNIIIGAAASTVPAKTSGENVITRTQDQIQIELLKDSKFDSTVTG-----NTLN 1784  
1160 KSRLL--DVTVTIGELHELADKAKLVSDHKPDADQIKQIRQPDPTUREKRYESNPVKHY 1216  
1785 TNGLTKEGSPVTKGINAGKKTINVDAGVNAKDAVN--KSQDNLAAKQNTDAAVK 1842  
1217 TDMGFTINKALEANYDAVKAFAKKEHGVNLTTRTVLESQSGAEAKLK-----NT 1271  
1843 YDDAKTKDKVTLKKGQ--TVLDNVKAGH--ISSTKEAVNGSQIHNSIKSIGGNT 1898  
1272 LLSLDSGESMFSRSYGGVSTV-----FVPTLSKKVPVPVPIGAGITLDRAYNLSPRT 1326  
1899 VVNPFG--SLTNNIGGTGKNINDAISEVKNATKAKTIVTEGDNIVVKTEN-----1950  
1327 SGGLNVSPRGDGVSGNIMVATGHDVM--PWTGKKSAGN-----ASDWLSAKH 1374  
1951 -----KDG--STNVEVSTKDLTNSVTTGDTVLNNGLTIKDGPSTIKDGVNAGG 1999  
1375 KISPLRIGA-----AVSGTLQGLTQNSLKFKLITELDEPGFIHGLTHGLTIPAEELQ 1427  
2000 KKITDVANGVTAQNSKDAVGAQVHHISNSIK-----NSIGGNT--VVNP-----2042  
1428 GIEHQMGQSKLITFSDVTSANLDRAGINLNEGSKPVGTVARSAGLSASANLAAGRE 1487  
2043 -----DGLSTNNIGGTGKNINDAIAK-SVDEKVTNGVNDLTQKGLNFGAN-----D 2088  
1488 RSTTSG-----QPGSTTSASNNRPTFNGVAGANL-----TAALGVAHSTHEGKFPV 1536  
2089 QKTTQKGAVHRKLGDT-----INIVG-GADAKTAEDKTSGENIITRTEDGVKIE 2137  
1537 IFP--APTSTNVSAALALDNRTSQSISLELKAEPVTSNDIS-----ELTSTL-GHKFKDS 1589  
2138 MLKDVKFSVNVG-----GHVNLNOGLIIGKGPSITVNGINAGKQITNVADGINAKDA 2191  
1590 ATTKMLAALKELDAPQBLHLOQHFSAKDVVDGEREYAVRNKLVIRQQAADSHSM 1649  
2192 VNKGLD--KQINEVK--DQI-----GKD-IGKLSDAHVOYDKD---KNGVNDKNSV 2235  
1650 ELGSASHSTYNNLSRINDGIVELHFKHFDALPASSAKRLGEMMNDPALKLIKQLQ 1709  
2236 TLGGGEGKGLNKVA---DG-----KVAEGSKDAVGGGLWNTQNVKNSNDIKNIQ 2285  
1710 ST-----PFSSASVSMELKDGHLREOTEKAILDKGVGREV 1744  
2286 NNIDNISGKAGLVQOQKENGELTVCKDTGGTINMASKEG--DRVVOGVKQGEI-----2338  
1745 GVLFDQRNNLRVKS---VSVSQSVKSEGFNT---PALLIGTNSAAMSERN-----IGT 1794

2339 -----XAGSNOAVNGGQIHKISSIKNSIGGNTTIDPKDSITNNIGGTGKNINDAIGT 2394  
1795 INFKYQDQNTPRRFTLEGGLAQANPQVASALTDL 1829  
2395 LN-----QSNQELGNKIINL 2409  
RESULT 6  
US-09-543-681A-4998  
; Sequence 4998, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4998  
; LENGTH: 1589  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4998  
Query Match 2.5%; Score 238; DB 4; Length 1589;  
Best Local Similarity 19.5%; Pred. No. 1.3e-07;  
Matches 341; Conservative 238; Mismatches 603; Indels 570; Gaps 89;  
289 SPLKPSQMLKSGAGVTPPLAVTLDKGKLQAPDPNPALNTLLKQTL---GKTOHYL-AH 344  
172 NPLFENGQLKGYSTLNNTNLLSLGKGL-----NTTGLDLIAPRIDSRGKITAAEISAF 226  
345 HASSDGSQHLHLLDNKGHLFDIKSTATSYSVLHNSHPGEIKG-----KLAQA 390  
227 TQNTFSQH-----FDLLSQKPVSAALDSVFFSGMQSGRIRIINTASGSGVKLAGK 277  
391 GTGSVSVVDGSKGKISLGSQTQ--SHNKTMLSQGEAHRSLLTGIWQHPAGAARP-----442  
278 FTANDLSVKNADNLTQDSQVRYDSYDK---DGSENVQNYRGGITVNNSSGSLTKTEL 333  
443 QGESIRL---HD-----DKIHILHPELGV-----WQ---SA 467  
334 KGKNTITVASSHNQIKASDLMGDDITLQGADLTIDGKLOQKETDIDNRWFYSWKYDVTK 393  
468 DKDTHSQLSRQADGK---LYALKDNRTLQNLSDKSEKLVDKIKSYSDVQRGQVAIIT 523  
394 EKEQIQIGSQIDAKNNATLTATKGDVTLDAKINAGNNLAINANKDIHING-----LI 447  
524 DTPGRHKMSIMPSLDASPEHSIS--LSLHPADAHQGLHKGSELBAQSAVISHGLRVVAD 581  
448 EKESR-----SENGNKRNTSHLESGSWNSHQTETLKASelta-----GKDLGLD 493  
582 SEGKLFSAIPKQDGNELKMKAMPQHALD--EHFGHDHQSISGFPHDDH-----GQNA 633  
494 AQGSITAGAKLHANENVL--VNAKDNINLVQKTNNDKTVT---DNHVMWGGIGGQGN--547  
634 LVKNNFRQ---HACPLGNDHQFHPGWNLTDAVLDNOLGLHHTNPEPPEILDMHGLSL 690  
548 ---KNNNNQOQVSHATQLTADQGL-----LLAADNNVNI-----578  
691 ALQEGKLYFDQLTKEGWTCAESDCKQKGLDGAAYLLKDGVEKLN--INQSTSSIKGT 749  
579 -----TGSQ-----VKNGQAFVKTTOGDVVIDNAMSETISKIDERT 615  
750 ENVFSLPHVRNKPFG--DALQGLNKDDKAQAMAVIG--VNKYLALTEKGD-----IRSF 800  
616 GTAFNITKSSHKNETKQSTGSELISDAQLTVVSGNDVNVNGLSKSADKLGISLGD 675  
801 QIKPGTQ--QLERPAQTLSREGISGELKD-----IHVDHKNQNYALTHEGEVHFHPRE 852

Db 676 NVKSAQVYKIDDEKTSLATGAKEVEDKQYAGFHITHTNKNSTSE-----724  
 QY 853 WONGAESWHLKALPQSEKLSLMSHSHKPIATFEDGSOHQLKAGWHAYAAPERG 912  
 Db 725 -----TEQANSTISGANVDLQANKNVIFA-GSDLKTTAGN-----758  
 QY 913 LAVGTSQSQTVF-----NRLMOGVKGVKIPGSGUTVKLSAQTGMGTGAEGKVSSEKFSERI 968  
 Db 759 --ASITGDNVAFVSTENKQKQTDNTDTTISGG-----FSYTGQVD-----KVGSKAD---802  
 QY 969 RAYAFNPTMTSTPRIPKAAVATO-----HGWQRECLKELYEMOGA 1009  
 Db 803 --FOYDKQHTQTEVTKNRSQTEVAGDLTITANKDLLHEGASHVEGR-----YQESG-853  
 QY 1010 LIKQDLAHVRNHPAPODLSKLETLDLG-EHGAELNDMKRPRDELEQSASTRSVTVLQ 1068  
 Db 854 -----ENIQHLAVNDSETSKTDSLNVGIDGVNL-----D 883  
 QY 1069 HGVLSKNGEINSEFPKSPGKALVQSFVNRSGQ---DLSKSLQAAVHATPPSAESKLQSM 1126  
 Db 884 YSGVTK-----PVKKAIEDGVNTTKPGNNTDLTKV-----TARDAIANLANL 926  
 QY 1127 LG-HFVSAGVMSHOKGEIPLGRQDPNDKALTAKSLDITVTIGELHELADKAKLVSD 1185  
 Db 927 SNLETPNVGVEIGKGG---GSQKSQDSQAVSTS-----INAGKIN-----965  
 QY 1186 HKPDADQIKQLRQPDITLREKRYESNPVHYTDMGFTHNKALEANDYKAFINAFKXEH 1245  
 Db 966 -----IDSNKHLH---DQG-TH-----YQSTQ 983  
 QY 1246 HGVNLTTRTVLESQSAELAKKLKNTLLSLDSEMSFSRSYGGVSTVFVPTLSKKVPV 1305  
 Db 984 EGISUTANT-----HTSEVAQDKHOT-----TFHETKGGG-QVGSTKGTGSDIT 1026  
 QY 1306 PVIPGAGITLDRAYNLSFRTSGGLNVSGFRDGGVSGNIMVATCHDVMPTMGTGKTSAGN 1365  
 Db 1027 VAIKEGQTTNNA---LMTKAKGSQFTS-----NGDISINVGEDA---HYEGAQFDAQ 1075  
 QY 1366 ASDWLSAKHKISPDLRIGAAVS---GTLQGTQLNSLKFKL-----TEDELPGFIHGLTH---1416  
 Db 1076 GKTVINA---GGDLTLTAQATDTHESESQNVNGSANKLVGTPPEKSDYGGGFNAGTTTHS 1131  
 QY 1417 -----GTLTPAELLQKGEIHKOMK-QGSKLTFPSVDTSAN---LDLRAGINLNEDGSK 1463  
 Db 1132 KEQTTAKVCAITGSGOIEINAGHNLTLQGTLSSEQDIALNATNKVDLQSA---SSERTEK 1189  
 QY 1464 PNGVTARVSAGLS-----ASANLAAGSRERSTTSGQFSTTSANNRPTFLN 1510  
 Db 1190 GNNLSGGVQAGFGKMTDDASSVNLGSAQFAIGKQDEKSVSREGGTTNNSGN-----1242  
 QY 1511 GVGAGANLTAALGVHASTH-EGKPVGIFPAFTSTNVSAALALDNR-TSQSISLELKRAE 1568  
 Db 1243 -----LTINGNSVHLOQAQVN-----SKQTQLTSQSGDIEITSQ 1277  
 QY 1569 PVTNDISELTST-LGKHFKDSATTKMLAALKELDDAKPAQLHL-----QQHF 1617  
 Db 1278 ---STDYKNWMTDIFGNGKXTNTP-----KEVTEKPAITSIHNIQKLLVNVEDQKT 1329  
 QY 1618 SAKDVWDERYEAARNLKLVI---RQQAADSHMELGSAHSSTYNNLSRINNDGIVELL 1675  
 Db 1330 SHQWATLETGTLTINSKDLTSGANVTADSVTGNVGSLSNIAQRESNRHVTVG-VNVG 1388  
 QY 1676 HKHFDALPASS-----AKRLGEMMNDPALKDI-----KQLQSTPPSSASVSMELK 1723  
 Db 1389 YNHTND-PRKSSQVKNKAKAGSLL---EKTIKOTIDSGIKSSTDAISDKYNSLSTIADK 1444  
 QY 1724 DGLBEQTEKALD---GRKV-----REEVGVLFDNRNLRKVSVSQSV 1765  
 Db 1445 TGISDET-KAKIDQFGKGVGKIKNIIVTGAEGHTANADIKVTHVD-NDAVTKTTLSTSN 1502  
 QY 1766 SKSEGFNTPALLTGTSNAAWSMERNTGTINFVKQ---DONTPRRTLEGG---IAQANP 1820

Db 1503 DLSLVNNGSTKLTAEG-----IASKQGOVDLGGSSVKLENIGHHVEAGADL 1549  
 QY 1821 QVASALTDLKKKE 1832  
 Db 1550 DLKSSVVDLAKQ 1561

RESULT 7  
 US-09-841-786-1  
 ; Sequence 1, Application US/09841786  
 ; Patent No. 6669940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAGARAJA, T. G.  
 ; APPLICANT: STEWART, GEORGE C.  
 ; APPLICANT: NARAYANAN, SANJEEV K.  
 ; APPLICANT: CHENGAPPA, M. M.  
 ; TITLE OF INVENTION: RECOMBINANT FUSOBACTERIUM NECROPHORUM LEUKOTOXIN  
 ; FILE REFERENCE: 30296  
 ; CURRENT APPLICATION NUMBER: US/09/841,786  
 ; PRIOR FILING DATE: 2001-04-24  
 ; PRIOR APPLICATION NUMBER: 09/558,257  
 ; PRIOR FILING DATE: 2000-04-25  
 ; NUMBER OF SEQ ID NOS: 15  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 3241  
 ; TYPE: PR1  
 ; ORGANISM: Fusobacterium necrophorum  
 US-09-841-786-1

Query Match 2.5%; Score 236; DB 4; Length 3241;  
 Best Local Similarity 18.9%; Pred. No. 5.6e-07;  
 Matches 416; Conservative 339; Mismatches 776; Indels 668; Gaps 117;

QY 9 EHKAHVHTAHPVGHVVALQOQSSSSPQAAAAAAGKNGKMPRIHQFSTADGTS 68  
 Db 155 DFKKALEAKHKGKVGNGIIPVDG-KVKIPLPNSGIVTEGK-----INAVEGIG 202

QY 69 --AAHQKKKFSL-----RGCLGTKKFSRS-----APOG 95  
 Db 203 LYAADIRLKDTAILKTGITDFKNLWISDRINSGLTDLKATKTKSGDILLSAHDSPK 262

QY 96 QPGTTHSGATLR-----DLLARDGETQHEAAPDAARLTTRSGVVRN-- 140  
 Db 263 AMGNSTVGGKIEEVKNTYANIESDAVLEADGNIKTSAKATNGRFRFKGGEKETYTP 322

QY 141 --MDDMAGRPMVKGS--GEDKVPTQQRHQLNFPQMRQTMLSKVAHPASANAGDLQ 196  
 Db 323 LSLSDVEASVRVNGKVGKNDITAEAK---NF--YDALVTYLA-----KH 365

QY 197 SPHPHPSHHEI-----KEEVPVGTSTKATTAHA---DRVEIAQEDDD 235  
 Db 366 SFSFVTGSIPIINLNGFLGLTSSKVVIGDKAVEATEGKANTHSYSGVRATWGAATSP 425

QY 236 SEFQOLHQRL-----AREENPPQPKLGATPISARFQPKLTAVAESVLE 282  
 Db 426 LKTNLYEKANGKLLSIGAGYISAKNSN-----VTIEGVKSKGRADITSSENTID 479

QY 283 ---GT---DTTOSPLKPSQMLKSGAGVTPPLAVTLQK-KLQAPDN---PPALNTL- 329  
 Db 480 ASVSVGTMRDSNKVAL---SVLVTGENKS--SVKIAGAKVSESTDDVNVVRSEAINSR 534

QY 330 --LKQTLGK-----DTQHYL-----AHHASSD-----G 350  
 Db 535 AAVKGGGLDSCNGVVAANISYNNASSRIDVDGYLHAKKRLNVEAHNITKNSVLQTSGLG 594

QY 351 SQHLLDN---KGHLFDIKSTATSYSLVHNSHPEIKGLAQAGTGVSVVD-----398  
 Db 595 TSKPMNDHVYESGHLKSILD-AIKQRFPGSDSVNBEIKNKLTLNLFVSGVSATIANHNSAS 653

QY 399 ---GSKGKISLGGTGTOSHNTMLSQPCGAHRSLTLTGHWHPAGAARQGSIRLHDDKIH 455

Db 654 VAIGESRLS--SGVEGSSVNRALNEAQLRATTSSG-----SVAVRKEERK 697  
QY 456 ILHPELVGWSADKOT--HSQISROADKLYALKDNRTLQNLSDNKSSEKLVDKI 508  
Db 698 KLIGNAAVFGYKNNASVTIADHAEL--VSEKIDINSENK-IEYKNSPKMAKSVIDKL 754  
QY 509 KSVSVQDQGVAILTDPGRHKMSIMPSLDASPEHSISLHFADAHQGLHGSSEL---565  
Db 755 E-----LTKRAFGETWT--PEYDPKDIESTIEKLL--NAFSEKLDCKPELLN 798  
QY 566 -EAQSVASHGRVLVADSEGLFSAALPKQGDGNELMKAMPOHALDEHGFHDHQSGLFF 624  
Db 799 GERWTILPQ-----TSKGTATETIANYVQGMKLEEK-LPK-----GFK 839  
QY 625 HDHGOQNALVKN--NFRQOHACPLGNDHGFHPCWNLTDALVIDNQLGLHHTNPEPEIL 682  
Db 840 AFSEG-LSGILIKETLNF--TGVGNYANFH--TFSS-----GANGE-RDVS 879  
QY 683 DMGHGLSLALQEGKHLVFDOLTKGWTGA-----ESDCKOLKK-----GLDGAAYLLK 729  
Db 880 SVGAVSWVQEN-----YSKVSUG-KGAKLAACKDLNKAINKAETVNLVGNIGLARST 934  
QY 730 DGEV--KRLNINQSTSS-----IKHGTENVS-----LPHVR--NKPEPDALQGL 771  
Db 935 SGSAVGKGLNVORSKNSAIVEAKEKAELSGENINADALNRLFHVAGSFGSGGNAINGM 994  
QY 772 NK-----DDKAQAMAVGVNKKYLALTEKGDIRSFOIKP-----GTQQLERPAQT 815  
Db 995 GSYSGGISKARVSDDEAYLKA-----NKKIALNSKNDTSVWNAAGSAGIGIKN-----1043  
QY 816 LSREGISGELKDIHVDHK-----ONLYALTHEGEVHFQPR-----EAWQNGAES--860  
Db 1044 -AAVGAVAVNDYDISNKASTEDNDEGQSKDKNDDEVTVTAESLEVDATKTGTINSIS 1102  
QY 861 ---SWHKIALPQSSSKLSLDMSEHKP-----IATPEDGSQHL-----897  
Db 1103 VAGGINKVGSPESEKPKS-----EERPEGFGKIGNKVDVSKNKITSDMSDLTEKINY 1157  
QY 898 -----KAGWHAYAA--PERGP-LAVGTSQSQTVENRL-----MOGVKGKVPISGL 941  
Db 1158 ISEGVKKAAGNIPSNVSHPTDKGPSFSLGASGSVS--FNNIKETSAVDGVKIN-LKGANK 1215  
QY 942 TVKLSAQTCGCTGABGRKVSCKFSE-----RIRAYAFNPTMS--TPRPKNA---986  
Db 1216 KVEVTSSSTFVGAWGSSALQWNIHIGSGNSISAGLAGAAVNNIQSKTSALVKNSDIR 1275  
QY 987 -----AYATOHGWQ--GREGLKPLYEMOGALIKOLDAHNVHNAPOPDLOSKLETIDL-1037  
Db 1276 NANKPKVNALSGGTQVAAGAGLEAVKESGGQSKSYLLGTSASINLVNNEVSAKSENNTVA 1335  
QY 1038 GEHGAELLN-DMKRFRDE-----LEQSATRSVTVLQGHQGVKNGEINSEFKPSPGKA 1090  
Db 1336 GESESQKMDVDVTAQOATQVTCALNLOKNGKNGTVGATVTVAKLNKNVNASI--SGGRY 1393  
QY 1091 LVQSNVNRSGQDLKSLSQQAHVATPPSAESKLSQMLGHFVSAGVDMHQKGEIPLGRQR 1150  
Db 1394 T-----NVNRPAD--AKALLATTQVTAATV-----TGTTISSGAGLGNVQCAVSNNK-I 1438  
QY 1151 DPNDKTALPKSLIILDTVTIGELHELADKAKULVSHKPDADQIKOLRQOQFDTL---REKR 1207  
Db 1439 DNDVEASVDKSSI-----EGANEINVIADKVGSSD-----LAKEYQALLNGKOKK 1484  
QY 1208 YESNPVKHYTDMGFTHNKALE-----ANYD-----AVKAFINAPK 1242  
Db 1485 YLEDRGINTNGYTYTKEQLEKAKKEGAVIVNAALSVAGTDSAGGVALAVNTVKNKFK 1544  
QY 1243 KEHHGVNLRATRTVLESQGSAAELAK-----LKNLTLSDSGSMSPFSRYGGGVSTVFVP 1297  
Db 1545 AELSGSN-----KEAGEDKIHAKHVNVVEAKSSTVVVNAASGLAISKDAFSGMGSGAWQ 1597  
QY 1298 TUSKVPVPVPIPGAGITLDRAYNLSFRSTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMT 1357  
Db 1598 DLSNDTIKVDKGR-ISAD-SLVNANNILGVNVA-----GTIAGLSLSTAVG-----1643

QY 1358 GKKTSAAGNSDWLSAKHKISPDLRIGAAVSGTILQTLQNSLKEKLTDELPFGHIGLTHG 1417  
Db 1644 -----AAFANNTLIHNK-----TSALITGTKNPFPFGKNTKVN-----VQALNDS 1682  
QY 1418 TLT-----PAELLOKQIEH--QMKOGSKLT-----FSVDTSANLDIR-----A 1453  
Db 1683 HITNVSAGGAASIKQAGIGMVSVNRGSDETEALVSDSEFEGVSSFNVDKQDKTINTIA 1742  
QY 1454 GINLNEGGKPNQVTVARVS--AGLSASANLAAGSRERSTTSG---QFGSTTSANNRPTFL 1509  
Db 1743 G---NANGGKAAGVATVAHTNIGKQSVIAIVKNSKITTTANDQDRKNINVTAKDYTTMT 1799  
QY 1510 NGVGAGANLTAALGVASHSTHEKPGVIPPAPFTSTVNSAALALD---NRTSQSISLELKR 1566  
Db 1800 IAVGVGAKGASVQGSASTTLNKTVS--SHVDQTDIDKDEENNGNKEKANVNVLAEN 1857  
QY 1567 AEPVTSN-----DISELTSTLGKHKFQDSATTKMLAALKELDDAKPAEQ 1609  
Db 1858 TSQVNTNATVLSGASQAAGVAVVANKITQNTSAHIKNSTQVNRNALVKS-----1908  
QY 1610 LHILQOHFSADV--VGDERVEAVRNKLKLVIRQOADSHEMELGSHSHTTYN-NLSRIN 1667  
Db 1909 ----KSHSSIKTIGIG-----AGVGAGAGVTVGSVAVNVKIV 1940  
QY 1668 NDGIVELLLHFKHFDALPASSAKRLGEMMNNDPALKDIIKOLQSTPFSS--ASVSME---1721  
Db 1941 NNTIAELNHAKITA-----KGNVGVITESDAVIANVAGTVSGVARAAIGASTSVNEITG 1994  
QY 1722 ----LXDG--LREOQTEKAI-LDGKVGREBVGVLFOQRNNLRV-KSVSVSQSVSKSEGF 1771  
Db 1995 STKAIYKDSVTIAKEETDDYITTGQVDK---VVDKVFKNLINEDLSQKRKISNKKGF 2050  
QY 1772 NTPALLGT-----SNSAAMSMERNICTINPK--YGQDQ 1803  
Db 2051 VTNSSATHLLKLLANAAGSGQAGVAGTVNINKVYGETE 2089

RESULT 8  
US-09-134-001C-5080  
; Sequence 5080, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ IDS: 5674  
; SEQ ID NO 5080  
; LENGTH: 3696  
; TYPE: PR1  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5080

Query Match 2.5%; Score 233; DB 4; Length 3696;  
Best Local Similarity 19.0%; Pred. No. 1,1e-06;  
Matches 351; Conservative 277; Mismatches 717; Indels 506; Gaps 86;

QY 274 TAVASVLEGDTTQSPKPKQSMKLGSGAGVTPPLAVTLIDKGLQLAPDN---PPALNTLL 330  
Db 53 TAQSDGGLNT--SQSNPISSEETNTLSGTQVPS---STENKQTVNPNHNAQPIAINT--106  
QY 331 KQTLGKDTQHYL---AHHASQSGSHLLLDNKGHLFDIKSTATSVLHNSHPGEIKGKL 387  
Db 107 EBAESAQTASYTNINENNDTSDGLHVNPQAKHIEAQSEDVTNHTNHSNSISIPENK-165  
QY 388 AQAGTGSVSDGKSGKISL--GSGTQSHNKTMLSQPGEAHRSLITIGIQHPAGARPQ-444

166 --ATTSSSKPKRGKRSGLTSGNDTSTTQNTDNLN---TG-----PNGI 209  
 445 ESIRLHDDKIHLHPGLVWQADKTHSOLSRQADGKLYALKDNRTLQNLSDNKSSEKL 504  
 210 NTVIFDD-----LGKTSNR-----SRPEVKV 233  
 505 VPKISYVDQGVAILPDTFGRHMSIMPSLSDSPESHISLSLHFA-----552  
 234 VDSLNGFTWVGKGVLLNSV--LERTSVFDSAD--PKNYQALDNVVALGRKGNPDNDH 289  
 553 DAHQGL-----LHGKSEL-----EAQSVASHG--RLVVDASGKLPFAIPKQDGG-----597  
 290 DGFNGIEKFEFVNPSEIIFSTNTAKNRKGGTQLVLENAENNOBIASDIQGGGVYRL 349  
 598 -----NELKMKAMPQHALDEHFGHDQI--SGF-----PHDDHG-----OLNA 633  
 350 FKLDPNVHLKQVFLPMEIHSDFRKIQOHGGRYYSFIDITGVNSGHLVVKRQVVK 409  
 634 LVKNNFROOHAQCLNDHQFHPGWLTDALVIDNQGLHHTHPEHEILD-----MG 685  
 410 NVKNGKEFEVNTRIENNGNF-----AAAIQONELTYKVTLPENFEYVDNSTEVSFVNG 462  
 686 HLGSLALQEGKLHYFDQLTGKWTGAESDCKQLKGLDGAAYLLKDEGVKRLNI-----738  
 463 NVPNSTVNPFSVN--FDRQNHLLTFSSNGLNLSRQADVARFLPN---KILNIRYKLRPV 517  
 739 NQST-----SSIKH-----GTENVFLPHVRNKPBPGLDGLQNKDD 775  
 518 NISPREVTENEAIKYKTFSEYIYNTNDVTGQOTPFPSINVIMK---DDLSEQVKNDI 574  
 776 KAQAMAVGVNKYLALETKGDIRSPQIKPGTOOLRPAQT-----LS 817  
 575 IPSNYTLASYNKYNKLERKAQTVLDEETNNTFPNQYSQTQIDDLHELQTLINRVAS 634  
 818 RE-----GISGELKDIHVHDKONLYALTHEGEV-----FHQP 849  
 635 REINDKAQEMTDVAVDSTELTTEEXDTLVDQIEN---HKNEISNIDDELTDGGERV 689  
 850 REAWONGAESSSWHKLALPQSESKLKS-----LDMSEHKPTATFEDGSG--HOLKA 899  
 690 KEAGHTLESITPHTVPFNARQVNNRQADQKTLIRNNE-----ATTEEQNEAIRQVEA 745  
 746 HSSDAIA-----KIGEAETDTTNEARDNGTKLIATDVPNTKKAERAAVNTSANSK 798  
 959 -----KVSSEKESIRIAYANPTMSTPRPKNAYATQHWQREGKLPYEMOGALIK 1012  
 799 IKDINNNTQATLDER-----NDALVNRSKDEAIQNTAQGNDDVTEAQNNGTNTIQ 852  
 1013 QLDAAHV--RHNA-----POPDLQSKLETLDLGEHGAELLNDMKRFRDEL---EQSATS 1063  
 853 QVPLTPVKRQAIATINAKADEQKLIQANNATTEKADAEKNEAVITANQNTINAT 912  
 1064 TVLGHQHVGLKNSGEINSEPKPS-----PGKALVQSFVNNRSGQDLKSLSQ---1109  
 913 TNRDVDOAQTTGSG-IISAIQSPATKIKEDARAAVEAKAIQNNQINSNNMATTEBEKDAL 971  
 1110 -----QAVHATPPSAESKLSQMLGHFVSAGVDMSHQKGE--IPLGRORDNDKTLATKS 1161  
 972 NQVEAHKQAAIATINQASTQOQ-----VSEAKNNGINTINQDQEN---AVKN 1016  
 1162 RLIIIDTVTIGELHELADKAKLVSDHKPDADQIKQLRQFQDTLREKRYESSNPVKHYTDMGF 1221  
 1017 ---NKTII--LEQKNEKKSAIAQTQDA-----TTEBKQEAIVSAVSOAVNTGI 1059  
 1222 THNKALEANYDAVAFINAPK--KEHHGVNLTFR-----TVLESQGSALA 1265  
 1060 THINQANSNDVDQELSNAEQIITHTNVVQKKPQARQALIAKTNEKQSAINSNEGTE 1119  
 1266 KKLXNTLLSDSGSMFSRSGGVSTVFVFTLSKKVPVPIQAGITLDRAYNLFSR 1325

1120 EKQK-AIOSLNDAKNLADBEQ-----ITQAASNQ-----NVDNALNIGSN 1158  
 1326 TSGGLNVFGRGGVSGNI-----MVAIGH-----DVMPYMTGKTKTSAGNASD 1368  
 1159 IS-KIQTNFTKKOQARDQVQKQFQKEAELNSTPHATQDEKQDALTRLTQAKETALNDIN 1217  
 1369 WLSAKHKISPDRLIGAAVSGTLQGLTQNSLKPFLTEDELPGLF--HGLTHGTULTPAELIQ 1426  
 1218 QAQTQNVDTALTSGIQ---NIONTVQNVVRKQKEAKTINDIVQHKQSIQNNDDATTEE 1274  
 1427 KQIEHQMGKSG--ITFSVDTSANLDLRAGINLINEDGSPKNGVTARVSAGLSASANLAAG 1484  
 1275 KEVANNLVNASQONVISKIDNATTNNQIDGI--VSDGRQOSINAITPDTSIKRNAKNDIDIK 1333  
 1485 SPERSITTSQFSGSTTSASNRRPTFLNGVCGAGANLTAALGVAHSST---HEGKFPV---1536  
 1334 AADKKIKIQRINDATDEEIQE---ANRKIEEAKIEAKONIQRNSTRDQVNEAKTINGINKI 1390  
 1537 --IPFAPTSTNVSAALALDNRTSQSI-----SLELKAEPVTSN 1573  
 1391 ENITFA--TTVKSEARQAVQNKANEQINHIQNTPDATNEKQEAIRNVSAELARVQAQIN- 1448  
 1574 DISELTSTLGRHFKDSATT---KMLAALKELDDAKPAQLHLIQ--QHFSKADVVGDERY 1628  
 1449 --AEHTTQGVTKIKDDAITSLSRINAQVVEKESARNAIEQKATQOTQFNNNDNATDEEK 1506  
 1629 EAVRNKLKLVIRQAAADSHSMELGSASHSTYNNLSRINN-----GIVELLHKHFD 1681  
 1507 EVANNL-VIATKQKSLDN-----INSLSS--NNDVENAKVAGINEIAN-----1546  
 1682 ALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKDLGRLQTEKAI--LDGKV 1739  
 1547 VLPATAVSKAK-KDIDQKLAQIINQIQTHTATT-----EKEAAIQLANQKS 1594  
 1740 GREEVGVLFQDRNNLRVKSVSQSVKSEGFNTPALLGTSSAAMERNIGTINFKY 1799  
 1595 NEARTAIQNEHSNN-----GVAQ--AKSNGIHEITELVMPDAHKKSDAKQ-----SIDNKY 1642  
 1800 QDQNT-----PRFTLEGGI-----AQANPQVASALTD 1828  
 1643 NEQSNITNTTPATDEEKQKALDKLAKDAGYNKVDQAQTNQOVSDAKTE 1693

RESULT 9

US-09-268-347-48  
 ; Sequence 48, Application US/09268347  
 ; Patent No. 6335182  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Loosmore, Sheena M.  
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS  
 ; FILE REFERENCE: 1038-860  
 ; CURRENT APPLICATION NUMBER: US/09/268,347  
 ; CURRENT FILING DATE: 1999-03-16  
 ; NUMBER OF SEQ ID NOS: 54  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 48  
 ; LENGTH: 2048  
 ; TYPE: PRT  
 ; ORGANISM: Haemophilus influenzae  
 US-09-268-347-48

Query Match 2.38; Score 216; DB 4; Length 2048;  
 Best Local Similarity 18.11; Pred. No. 7e-06;  
 Matches 311; Conservative 244; Mismatches 612; Indels 549; Gaps 78;

Qy 392 TGSVSDVG-----KSGK-ISLGGTQSHNKTMLSQPG---EAHRSLLTGIWQH 435  
 Db 86 SGTAQADGDRALIAIGENANAGGQAIAGSNKTVNGSSLDKIGTDATGQESIAIG----141  
 Qy 436 PAGNARPQGE-SIRLHDKTHILHPELGVWQSAADKOTHSQLSROADGKLY--ALKDNRTL 492  
 Db 142 --GDVKASGDASIAIGSDLLHL-----DQHG--PKHPKGTLLINDLINGHAVL 186

493 QNLSDNKSEKLVKIKSYSDVQRGOVAILITDFGRHKMSIMPSLDASPEHSISLSLHFA 552  
 187 KEIRSSKDNV---KYRRITASGHASTAVGAMSYAQGHFSNAGTRATAKSAVSLAVGLA 243  
 553 DAHQG---LHKGSELEAQSVAISHGRVLVVADESGK--LPSAAIPKQGDGNEUKMKAMPO 607  
 244 ATAEGQSTIAIGSDATSSSLGALAGAGTRAQIQGSIALGQGSVVTQSDNNRSPAYTPNT 303  
 608 HALDEFGHDHQSIFGPHDDHCOLNALVKNFPQOQHACPLGNDHQHPGWNLTDALVI-- 665  
 304 QALDPKFOATNNTKA-----GPLS---IGSNSIKRKLIINGV-----AGVNTKIDAVNVAQ 349  
 666 -----DNQGLGHTNPEPHEILDGMHGLSIALQEGKHLVFDQL 703  
 350 LEAVVWAKERRITFOGDDNSTDVKIGLNT-----LTIKGA--ETNAL 392  
 704 TKGWTA--BSCDKQLKGLDGAAYLLKQGEVRLNINQSTSSIKHG-----TENVPSLP 756  
 393 TDNNIGVVKADNSGLKVKL--AKTLNLTENVNTTLN--ATTIVKVGSSSTTAEELSDS 449  
 757 HVNKEPEGD-----ALQGLNKDKQAAMAVIGVKNYLALTEKDIRSFQIKPGCTQQ 808  
 450 LTTQPNQTSQSTKTVGVNGVKTFTNNAETTAIGTR----- 488  
 809 LERPAQTLREGISGELKDIHVDHKQNYALTHEGEVHFQPREAWQNGAESSMHKLALP 868  
 489 -----ITRDKI--GFARDGDVDEKAPY-----LD 511  
 869 QSSKLSKLSMSHEHKEPIATPEDQSOHQKAGGHVAAAPERGLAVGTSGSQTVFRLM 928  
 512 KKQLKVGSAI-----ITDNG--IDAGN-----KKISNLAKGSSANDAVTIEQL 553  
 929 QGVKGVIPQSGTLV-----KLSAQTGMT-----CAEGRKVSSKFSERIRA 970  
 554 KAAKPTLNAGAGISVTPTEISVDKSGNVTAFTNIGVKTIELNSDG--TSDKES--VKG 609  
 971 YAFNPTMTSTPRPI-----KNAAYATQH 992  
 610 SGTNNSLVTAHSLASYLENVNRTADSALQSFTVKEEDDDANAITVAKDTTKNAGAVSIL 669  
 993 GWQREGKLPYEMOGALIKOL-----DAHVNRHNAQPDLOS---KLET 1034  
 670 KLGKNGLTATKKGDTVTGFLSQDSGLTIGKSTLNNDGLTVKDTNEQIQVGANGIKPTN 729  
 1035 LDLGHEGAEGLNDMKRFEDELEQASRATSVTVLQHQGVKNGEINSFKEPSGKALVQS 1094  
 730 VNGNPGGIGANTAKITRDKLGFAGS-----DGAVDTNKPYLQDKLQGVNVTN 780  
 1095 FVNRSGQDLSKLSQQAQVHATPPSAESKLQSMGLHFVSAGVDMHQKGEIPLGRQRPND 1154  
 781 TGINAGGKAIT-----GLSPTLPSI-----ADQSSR--NIBLGNTIQDKD 818  
 1155 KTAITKSLIIDL-----VTIGELHELADK-AKLIVSD- 1185  
 819 KSNAAASINDILNTGFNLKNNNNPIDVFTYDIPFANGNATTAIVTHTDANKTSKVVYDV 878  
 1186 -----HKPDADQIKOLRQOPDITLREKRYESNPVKHYTDMGFTHNKALE---ANYDAVK 1235  
 879 NVDDTIHLTGDDNKKLGKVTTKLNKTSANGN-----TAINFVNVSSDEADALVNAKDIA 933  
 1236 AFINAFKBBHGWNLTTRTVLESQGSABELAKKLKNTLISLDSGSEMSFSRSGGVSTVF 1295  
 934 ENLNTLAKEIHTTKGTADTALQT---FTVKVDENNADANAITVQCKNANNQVNTL- 988  
 1296 VPTLSKKVPVPIPGAGITLDRAYNLSPS-RTSGGLNV--SPGRDGGVS-----GNIMVA 1347  
 989 ---TLKGE-----NGLNITKDKNGVTFGINTSGLKAGKSTLNDGGLSIKNPTGSEQIQ 1040  
 1348 TGHDMVPMYMGKTKSAGNASDWLSAKHKISIPDLRIGAAVSGTFLQGLNSLKFKLTEDBL 1407  
 1041 VGADGVKFA-----KYNNGVVGAGIDGTT-----RITRDEI 1072  
 1408 PGFIHGLTHGTLTPAB--LLOKIEHQMKQSGKLTFSVDVTSANL---DLRAGINLNEDGS 1462

1073 -GFTG--TWGSLDKSPHLSKGIN---AGGKIT-----NIQGEIQANSHDAVGG 1119  
 1463 KPNQVTARYSAGLSASANLAAGSRERSTTSQGFSTTSASNNRPTF-----L 1509  
 1120 KIYDLKTELENKISSHTAKTAQNSLHEFSVADEQGNFTVSNPYSSYDTSKTSVDITPAGE 1179  
 1510 NGVAGANITAAALGVASHSTHEKPGVIGIPAFITSTVWSAALDNRKTSQISLELKRAEP 1569  
 1180 NGITTKVN---KGVVRVGDIDTKGLTTPKLTIVGNNGKIVIDSONGQ----- 1224  
 1570 VTSNDISELTSTLGHFKQSA---TTKMLAALKELDDAKPAEOLHILQOHFS-----AK 1620  
 1225 ---NTITGUSNTLVANTVNDKSGVRTTEQGNIIKDEKTRAASIVDVLSAGFNQNGEAV 1281  
 1621 DVVGDERYEAVRNKKLVIRQQAADSHMSLGASHSTTYNNLSR-----INNDGIVE 1673  
 1282 DFV--STYDIV-----NFDAGN---ATTAKVTYDDTSKTSKVYDVNDTDTIE 1324  
 1674 LLHKHF-----DAALPASS--AKELGEMMNDPALKDIIK 1706  
 1325 VKDKLGVKTTTLTGTGANKFALSNAQATGDALVKAASDIVAHSNLTLSGDIQTAKGASQ 1384  
 1707 QLQSTPSSASVS-----MELKDGLEOTEKAILDQKVGREEV--GVLFQDR 1751  
 1385 ANNSAGYVDADGNKIVYDSTDNKYQAKNDGTVDKTEKAKOKLVAAQATPDGTGLAQ-- 1442  
 1752 NNLRVKSVSVSQSV---SKSEGFNTF--ALLLGTNSAAMSERN-----IGTIN---F 1797  
 1443 --MNVKSVINKEQVNDANKQGINEDNAFVKLEKAAASDNKTKNAAVTVGSLNAVAOTPL 1500  
 1798 KYGQDQNTPRR-----FTLEGGIAQANPQVASALTD 1828  
 1501 TFGADTGTAKLGETLTIKGGQDNTN-----KLTD 1531

RESULT 10

US-09-489-039A-7973  
 ; Sequence 7973, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 7973  
 ; LENGTH: 2680  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 ; US-09-489-039A-7973

Query Match 2.2%; Score 211.5; DB 4; Length 2680;  
 Best Local Similarity 18.2%; Pred. No. 2.2e-05;  
 Matches 365; Conservative 255; Mismatches 709; Indels 675; Gaps 81;  
 QY 188 ANAGDRLOHSPHPHPSHHEIKKEPVGSTSKATTAHADRE---IAQEDDDSEFQQLHQ 243  
 DB 532 AKPGDRLOHLPFSGISQSRITQAVNRRQITVTHYSETPRECVWAVESNDLFLO---Q 588  
 QY 244 ORLARERPPPPKGLGVATPISARFQPKLTAFAESVLGTTTQSPKPKQSMKLGSGAG 303  
 DB 589 YRVTGVKEN-----GDATLTITGVSHDPPKFAIDTGAIID 624  
 QY 304 VTPPLAVTLDKGKLQAPDNP---ALNTLLKQTLGKXDTQHY-----LAHHASSDG 350  
 DB 625 QRPVSV-----LPAGNQSPDDIVITSRVNVQGISVETMNVNWSVGAIAEYEAQ--- 675  
 QY 351 SQHLLLDNKGHLFDI-KSTATSYSVLHNSHPGKIGKLAQAQAGTGSYSVDG-KSGKISLGS 408

Db 676 ---WRNDGWINPRSSSTTGF---EVSG--IYAGKYLVRVRAINAAETSSG- 719  
QY 409 GTQSHNKTMLSPQGE-----AHRSLLTGI---WOHPAGAARPOGESIR----- 448  
Db 720 WAYSEKTLTGKVGEPFLALATSLVHGQVSVNEFFTGS-----GDTLRTLOYSKNOD 775  
QY 449 ---LHD-----DKIHLHPELG----- 462  
Db 776 GSAPMPLSDVAPYKGSYOQGLSMGAEPFYRARLVDRLGNESFWTCWQGMASDNFDDYY 835  
QY 463 ---VW-----QSADKTHSOLRQA----- 479  
Db 836 ENLTDAIKDTAAWERTORTISTQGIINTQOELEQTAEALKEAEDQAKQVSDIDASA 895  
QY 480 ---DGKLYALKDNRTLQNLSDNKS-----SEKLVDKIKGSYVDQRCQVAI----- 521  
Db 896 KSITADVDBGKISAVNKTITDEITSVNEALDSGLAQANKGVQEAASAVADANKQIATVNS 955  
QY 522 LTDTPGRHKMSIMPSLDASPEHSISLHFADAHQGLHGKSELBAQSAVISHGLVAVD 581  
Db 956 LTDSTQVRQSV---TDTAAEAENATIDLEIARVSKTLADGDAALNAQIKTAENGLKQSL 1012  
QY 582 SEGKLFSAAIKQGDGNELKMKAMPQHALLDHFHGHQHSIGFFHDDHGHQALVKNFRQ 641  
Db 1013 QVNTLTNAVQKQETADRIADNAKASQAADLELLAATQGEIASIESLTQVMKTDENLARE 1072  
QY 642 QHACPLGNDHQFHPGWNLTDAVIDNQLGLHHTNPEPHEILLDMHGLSLAQEGKLVHFD 701  
Db 1073 MSSLAAGANICF-----DSQVI-----WHFNN 1094  
QY 702 QLTGKWTGAESDCKQLKKG-----LDGAAY---LLK----- 729  
Db 1095 QTTGKWTGAGVPGVSODGWLRLPADSATDPYITSPGLAVDGAAYRFTMLFRKTKGKPV 1154  
QY 730 DGEVKRLNINOSTSIIKHGTENVPSLPHVRNKPBGDQALQKNDKQAQAAVIGVKYL 789  
Db 1155 AGEIRWVSAGENFNNTKR-----YIVAEPEYADGVATLTVRDIPWTKNIDRIIR--L 1203  
QY 790 ALTEKGDIRSFQIKGTQQLERPAQTLSREGISGELKHVDHQNLYALTHEGEVPHQ 849  
Db 1204 DLTNQODASNF-IEFDVIAGRPAGAS---TAALQDVRSTLSN---ALTAEA----- 1249  
QY 850 REAWQNGAESSSWHKALPOSEKLS-LMSHEHKPIATFEDQSHQLKAGGWHAYAAP 908  
Db 1250 ---QARSTLAAQMRGSDYDGSLEKVTSGLLYQEKARTARVTAISAE 1290  
QY 909 ERGPLAVTSGSOTVFNRLMGVKGKVIKPGSLTVKLSAQT---GGMTEAGRKVS----- 961  
Db 1291 VK-----ARESLQTFNDNKAASVSEL---SSLTQESQASRIRGGLETSLGKKADAAL 1342  
QY 962 SKFSERIRAYAFNPTM--STPRPIKNAAYATQHWQGR---EGLKPLYEMOQALIKOLD 1015  
Db 1343 TSLTKVEQOQATLTSQGAALTSLTNRVQGTETGLAGTNEALSGLQSVTVQGDRI----- 1398  
QY 1016 AHNVRNAPQDQSLQKLETLDLGHGAELLNDMKRFRDELEQSATRSVTVLGQHQGVLS 1075  
Db 1399 ---TSQGSITKL-TSDLGTNAAL-----AKKAEAAVATLTQO---VEQ 1437  
QY 1076 NGEINSEFKSPGKALVQSFNVRNSGQDLSKSLQAVHATPPSAESKLSMLGHFVSAGV 1135  
Db 1438 NGR-----DIRSNTDSITSLNQLVNGPQNRWSRLYP-----V 1471  
QY 1136 DMHQKGEIPLGRQRPNDKLTALTKSLRLDITVTIGELHELADKAKLVSDHKPAD----- 1191  
Db 1472 QLAN-AGTVP-----SFSDVRAVAPTV---VDEVADAAL--DFTSAGSVLIA 1513  
QY 1192 ---QIKQLRQOQDFTLREKYESNPVKHVYDMDG-----THNKALEANY 1231  
Db 1514 LYSQCVKVAADTTITL-----APGARVFDDTGAIFVNGQVAVGNASWNTVSPFELKAGW 1567  
QY 1232 DAVKAFINAFKBBHH---GVNLTR-----TVLES 1258

Db 1568 NTVEFLVNWQTCQAYINLGLKLDKVAEMYSGLVSAALANAAGVLSSNVSYQIGNEVVNS 1627  
QY 1259 QGSAAELAKKIKNTLTLSDSGESMSFSRSGGVY-----STVEVPILSKKVPVPIV 1309  
Db 1628 QSIOTLRNALTOTDANVASKADQOTMNSLTGRVEKTESGLTAANANITSLKSAVRAGNAS 1687  
QY 1310 GAGI---TLDRAVN--LSFSRST-----GGLNVSPGRDGVSGNIMVATGHVD 1352  
Db 1688 GGDLPNPTFPAYDQMGFSVVSTTAEEVPPSCPFGYAAARIASRDHHPFAFPATLNDV 1747  
QY 1353 MPYWTGKTSAGNASDWLSAKHKISPDRLRIGAAV-----SGTQGTQLNSLKFPLTE 1404  
Db 1748 IEISALVACGAGTANFNLYVGTAVRPDTSTGAPLMAGGKSPSATWQRT---TWRFKVTQ 1804  
QY 1405 DELP-GFI-----HGLTHGT-----LTPAELLQKGIHQMKQSKLTFSDVT 1445  
Db 1805 AMVDGXYIRPFIQISQNSPYGTVFWFVTDWHRMNVTAARQKQVDTADATAAADDLSLTTVTQ 1864  
QY 1446 SANLDRAGINLINEDGSKPENGVTARVSAGLSANLAAGSRERSSTSGQFGSTTSASNNR 1505  
Db 1865 QGNLLTSTG---NRTTQLENGU-ATTNAAVAKAD-ATAVQDLTNTVTQLGNDLTAANS 1919  
QY 1506 PTFLENG-----VGAGANLTAALGVASHSTHECKPV----- 1535  
Db 1920 ITKLTGNLANTDKALAKADATATLDTKVTQOGKTLESQNSLTLNLSLSQVAADD 1979  
QY 1536 --GIFPAFTSTNVSAALADNRTSQSISLELKRAEPVTSNDISELTSTLGHKFKDSATTK 1593  
Db 1980 ASGOIPGNLVNPSFERGLDGTGRS-----TATSVVEV-----SAPHS 2018  
QY 1594 MLAALK-ELDDAKPAEOLHILQOHFSADKVVGDREYEAVRNLKLVIRQQAADSHSMELG 1652  
Db 2019 GTRALKVDPCGSVPQYIPFVQ-----GRTELGWVVK-----EPG 2054  
QY 1653 SASHTYNNLSRINNDGIVELLHKHFD-----AALPASSAKEL- 1691  
Db 2055 ATTONGAGNKLKRICNSAGQVFERPYNSGVTGNTWTLVSGRWKATETASLPVTLNLYI 2114  
QY 1692 -GEMMNNDPALKDILIKQLQSTPFPSSASVSMELKDGLREOTEKAILDGKVGREVGVLFD 1750  
Db 2115 SGRYFDDFYVTVTDVDRVIDATAGAVTG-----LTSRVSTAEGAITSQS 2159  
QY 1751 -----RNNLRKVSVSQS-----VSKSEGTPTALLLGTSSAAMSMBERNI 1792  
Db 2160 QQLTNLQNSLTNNSVSKADATLTSVDNRVTEASGKLTQSQQLTNLANVLATRNA 2219  
QY 1793 GTINFKYQDQNTPRRTFLEGGIA 1816  
Db 2220 G-----DNLIPNDFLQGSTA 2235

## RESULT 11

US-08-409-995-4  
; Sequence 4, Application US/08409995  
; Patent No. 5646259  
; GENERAL INFORMATION:  
; APPLICANT: Barentkamp, Stephen I.  
; APPLICANT: St. Geme III, Joseph W.  
; TITLE OF INVENTION: Haemophilus Adhesion Proteins  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/409,995

FILING DATE: 24-MAR-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-61053/RTF

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1912 amino acids

TYPE: amino acid

STRANDEDNESS: double

TOPOLOGY: unknown

US-08-409-995-4

Query Match 2.24; Score 209.5; DB 1; Length 1912;

Best Local Similarity 18.8%; Pred. No. 1.8e-05;

Matches 346; Conservative 221; Mismatches 655; Indels 615; Gaps 84;

QY 236 SEFQQLHQOARERENPPQPKLGVAIPISARFPQKLTAVASVLEGTDTTOSPLKQ 295  
 DB 20 SELTRTHTKEL-RNRGDPV-----LATLLFATVQANATDEDEE-LDPVVRTAPVLSFHS 71  
 QY 296 MLKSG-----AGVPL-AVTLDKG---KLQAPD---NPPALNTLLKQ 332  
 DB 72 DKEGTGEKTEVNSNWGIYFDNKGVLKAGAITLKAGDNLKKXKQXTDEXTNASSFTYSLKK 131  
 QY 333 TLGKDTQHYLAHHASSDGSQHLLDNKGHLFDIKSTATSYSVLHNSHPGEIKGKLAQGT 392  
 DB 132 DLTDLT-----SVATEKLSFGANGDKVDITSDANGL-----KLAKTGN 169  
 QY 393 GSVSVSDG-----KSGKISLGSQGTQSH-NKTMLSQGEAHRSLTLGIWHQHPAGAAR 441  
 DB 170 GNVHLNGLDSTLPDAVNTGVLSSSSFTPNDEVKTRAA---TVKDVNLNAGW-NIKGAKT 224  
 QY 442 PQG--ESIRLHD--DKHILHPELGVQWSADKT-----HSQLSR 477  
 DB 225 AGNVESVDLVSAYNNVEFI-----TGDKNTLDVVLVTAKENKXKTEVFTPKTSVIK 276  
 QY 478 QADGKLVAKDNRTLQNLSDNKSSEKLVDKISYSDQGVQVAILTDTGPRHMSIMPSL 537  
 DB 277 EKGKLTGKENDTNKVTNATDN-----TD-EGNGLVTAKAVI 316  
 QY 538 DASPESHISLHFADAHQGLHKGKSELEAQSVAISHGRLVADSEGLFSAALPKQGDG 597  
 DB 317 DAVNKAGWRVKTITANGQNG-----DFATVASGNTVTFES-GDGTASVTKDTNG 365  
 QY 598 NELXKAMPQHALDEHFGHDHQISQFFHDDHQLNALVKNFRQOHACPLGNDHQFHPGW 657  
 DB 366 NGITVK-----YDAKVGDLKDFD--- 383  
 QY 658 NLTDALVIDNQLGHLHTNPPEHIL-----DMGHLGSLALQEGKLYFDOLTKGWTG-A 710  
 DB 384 ---SKKIVADTALTITVGGKVAETAKEDDKKLVNAGDLVTALGNL-----SWKAKA 433  
 QY 711 ESDCKQLKGLDGAAYLLKQGEV-----KRLNINOS-----TSSIKHGTENVFSL 755  
 DB 434 EADTDGALGIS-----KQGEVKAGETVTFKAGKNLVKQDGANFTYSLQDALTGLTSI 487  
 QY 756 PHVRNKPEGDALQINGKDDKAQAMA-----VIGNVKYLALTEKGDIRSFQIKPGTQQLER 811  
 DB 488 TLGGTNGNDAKTIVNKDGLTITPAGNGTGTGNT-LSVYKDG-----IKAGNKAITN 540  
 QY 812 PAQTL-----SREGISBELKDI--HVD-----HKQNLVYALTHEGEVFFHQF 849  
 DB 541 VASGLRAYDDANFDVNLNNSATDLNRHVEDAYKGLLENKANKQPLVTDSTAATVGDLR 600  
 QY 850 REAW-----QNGAESSS-----WHKLAIPQSESLSKLSLM 879

DB 601 KLGVVSTKNGTKESNQVQKQADEVLFTCAGATVTSKSENGKHTITVSAETK---ADC 657  
 QY 880 SHEHKPIATFEDGSQHLKAGWGHAYAAPERGLAVGTSGSTVFNRLMQGVKGKVI PGS 939  
 DB 658 GLE-----KQDGTIKLV-----DNQNTDNLVTVGNNGTAVTKG 691  
 QY 940 GL-TVKLSAQGTGTMGAERKVSSEKESERIRAYAFNPTMSTPRPIKNAVAATQHGQGRE 998  
 DB 692 GPETVKT-----GATDADRQKVT-----VKDATANDAKKVATVKDVATAINSA- 735  
 QY 999 GLKPLYEMOGALIKOLDANVRHNAPOPLOS-----KLETLDLGEHGAELNDMKRF 1051  
 DB 736 ---ATFVKTENLTSSIDEDNPTDNGKDALKAGDTLTFKAGKNLVKVRQDKNITFDLAKN 792  
 QY 1052 RDELEQSATRSVTVLQHQGVKLSNGEINSEFKPSPGKALVQGFVNNRSQDL----- 1104  
 DB 793 LEVKTAKVSDTLTIGN-----TPTGGTTATPKVNTSTADGLNFAKETA 837  
 QY 1105 ---SKSLQ---AVHATPPSAESKLOSMGLGHFVSAGVDMSHQKGEIPL----- 1146  
 DB 838 DASGSKNVYLKGIATTLTTEPSAGAK-----SSHVDLNVDAITKKSNAASIEDVLRAGWNIQ 892  
 QY 1147 --GRQD-----PNDKTALTKSRLILDVTI-----GELHELADKAKLVSDHK 1187  
 DB 893 GNGNVYVATYTVNFTDDSTGTT-----TVTVTKADGKGVKVIKAKTSVVKDHN 945  
 QY 1188 PDADQIKQLRQOQDFTLREKRYESNPVKHYTDMGFTHNKALEANVDAKAFINAFKE--- 1244  
 DB 946 GKLTGKOLKD-----ANNGATVSEDDG---KDTGTGLVTAKTVIDAVNKGSMR 991  
 QY 1245 --HHGNLTTRTVLESQGSALAKKLVNTLLSDSGESMSFSRSYGGVSVTFVPTLSKK 1302  
 DB 992 VTGEGATAETGATAVNAGNAE-----TVTSGTSVNFK---NGNATVATV- 1032  
 QY 1303 VPVPVPGAGITLDRAYNLFSRSTGGLNVSFORDGVSGVGNIMVATGHDVMP-----YMT 1357  
 DB 1033 -----SKDNGNINVKY--DVNVGDLKIGDDKKIVADTTTLTVT 1069  
 QY 1358 GKKTSGNAGSDWL-SAKHKISPDRLRIGAAVSGTQGTQNTLQNSLKFKLTEDELPGFHGLTHG 1417  
 DB 1070 GGVSVVPAGANSVNNKKL-----VNAEGLATALNLSWTAKADK---YADGESEG 1117  
 QY 1418 TLTPAEALLQKGIHQKQSGKLTFSVDTSANLDRAGINLNEGSK-----PNGVTAR 1470  
 DB 1118 E-----TQDEVKAGDKVTF-----KAGKNLVKQSEKDFYTSLODTLTGL 1157  
 QY 1471 VSAGLSANLAAGRSRSTSGQFSTTSASNNRPTFLNGVAGAGANLTAALGVAHSSHT 1530  
 DB 1158 TSITLGGTAN--GRNDTGTIVNKDGLTITLAN-----GAAAGTDASNGNTISVTK 1205  
 QY 1531 EGPVGIIPAPTSTNVSAAL-----ALNRTSQSISLELKAEPV-----TSNDISE 1577  
 DB 1206 DGISAG--NKEITNVKSALTKYKDTONTADETQKFEHAAVKNANEVFEVKGKATVSA 1262  
 QY 1578 LTSILGKHFK--DSATTKMLAALKELDDAK-----PAEQHLILQOHFS 1618  
 DB 1263 KTDNNGKHTVTDVAERKVGGLKEDTDGKLVKVDNTDGNLLTVDAKASVAKGEFN 1322  
 QY 1619 AKDVVGVDEREAVRNK---KLVIROQAADSHSMELGSASSHSTYNNLSRINNDGIVELL 1675  
 DB 1323 A--VTTDATTAGTINANERGVVVK-----GNGATATETDKKVTATV----- 1364  
 QY 1676 HKHFDAALPASAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKDGUREOTEKAIL 1735  
 DB 1365 ---DVAKAINDAATFVKVENDDSATID-----DSPTDDGANDALKAKDTL---TLKAK 1412  
 QY 1736 DGKVBREVGVLFOQRNRLRVKSVSQSKSEGFNTPALLLIGTSNAAMSERNIGTI 1795  
 DB 1413 NLKVRQDKNITFALANDLSVKSATVSKLS-----LGT-NGNKVNITSDTKGL 1460  
 QY 1796 NF-----KYGQDQNTPRFTLEGGTAQANPQVASALTD 1828  
 DB 1461 NFAKSKTGDDAN---IHLNG-----IATSLTD 1485



## RESULT 12

US-08-685-467-4  
 ; Sequence 4, Application US/08685467  
 ; Patent No. 6060059  
 ; GENERAL INFORMATION:  
 ; APPLICANT: St. Gene III, Joseph W.  
 ; APPLICANT: Barenkamp, Stephen J.  
 ; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
 ; STREET: Four Embarcadero Center, Suite 3400  
 ; CITY: San Francisco  
 ; STATE: California  
 ; COUNTRY: United States  
 ; ZIP: 94111-4187  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/685,467  
 ; FILING DATE: 22-JUL-1996  
 ; CLASSIFICATION: 424  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/409,995  
 ; FILING DATE: 24-MAR-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Silva, Robin M.  
 ; REGISTRATION NUMBER: 38,304  
 ; REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 781-1989  
 ; TELEFAX: (415) 398-3249  
 ; TELEX: 910 277299  
 ; INFORMATION FOR SEQ ID NO: 4:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1912 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: unknown  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: protein  
 ; US-08-685-467-4

Query Match 2.2%; Score 209.5; DB 3; Length 1912;  
 Best Local Similarity 18.8%; Pred. No. 1.8e-05;  
 Matches 346; Conservative 221; Mismatches 655; Indels 615; Gaps 84;

Qy	236	SEFQQLHQRLARENPQPPLKGLVATPIISARFQPKLTAVAESVLEGTDTTQSPKLPQS	295
Db	20	SELTRTHTKRL-RNRGDPV-----LATLLFATVQANATDEDEB-LDPVVRTPVLSFHS	71
Qy	296	MLKSG-----AGVTPL-AVTLQK-----KQLAPD-----NPPALNTLLKQ	332
Db	72	DXEGTGEKEVTENSWGIFYFDNKGVLKAGATTLKAGDNLKXKQXTDEXTNASSFTYSLKX	131
Qy	333	TLGKDTQHYLAHASSDGSQHLLDNKGHLDDIKSTATSYSLVLSHPGEIKGKLAQAGT	392
Db	132	DLTDLT-----SVATEKLSFGANGDKVDITSDANGL-----KLAKTGN	169
Qy	393	GSVSVVDG-----KSGKISLGSQTQSH-NKTMLSQPGEAHRSLLTGIWOHPAGAAR	441
Db	170	GNVHLNGLDSTLPDAVINTGVLSSTSTFPNDVEKTRAA-----TKVDVLNAGW-NIKAKT	224
Qy	442	PQG--ESIRLHD--DKIHLHPELGVWQSDAKDT-----HSQLSR	477
Db	225	AGNVESVDVLSAYNNVEFI-----TGDKNLDVVLVLTAKENKXKTKTEVKFTPKTSVIK	276
Qy	478	QADQKLYALDKNRTLQNLSONKSEKLVKDKISYSDVQRGQVAILLTPGRHKMSIMPSL	537

Db	277	EKDGKLTGKENNDTNKVTSNTATDN-----TD-EGNGLVTAKAVI	316
Qy	538	DASPESHISLSLHFADAHQGLLHGKSELERAQSVASHGRILVADSEGLFSAIPKQGDG	597
Db	317	DAVNKAGWVKTITTTANGQNG-----DFATVASGTNVTPEF-GDGTASVTQDNG	365
Qy	598	NELKMKAMPQHALDEHFGHDHQISGFFHDDHGQNLNVKNPFQHQACPLGNDHQHPGW	657
Db	366	NGITVK-----YDAKVGDLGKFD---	383
Qy	658	NLTDALVIDNQLGLHHTNPEPHEIL-----DMGHLSGLALQEGKLHYFDQLTKGWTG-A	710
Db	384	--SDKKIVADTTALTVTGKVAEIAKEDDKKLVNAGDLVLTALGNL-----SWAKA	433
Qy	711	ESDCKQLKKGLDGAAYLLKQGEV-----KRLNINQS-----TSIKHGTENVFSL	755
Db	434	EADTDGALEGIS-----KQDEVKAGETVTFKAGKNLKVQDGFANFTYSLQDALTLGTSI	487
Qy	756	PHVRNKEPQDALQGLNKDDKAQAMA-----VLGVNKLALTEKGDIRSFOIKEGTQQLER	811
Db	488	TLGGTTNGNDAKTVINKDGLTITPAGNGTTGTNT-ISTVKDG-----IKAGNKAITN	540
Qy	812	PAQTL-----SREGISGELKDI--HVD-----HKQNLVYALTHEGEVFPQP	849
Db	541	VASGLRAYDDANFDVLNNSATDLNRHVEDAYKGLLNLNEKNANKQPLVTDSTAATVGDLR	600
Qy	850	REAW-----QNGAESSS-----WHKLALPQSSSKLSKSLDM	879
Db	601	KLGWVSTKNGTKESNQVQADEVLFTGAGAATVTSKSENGKHTITVSVAETK---ADC	657
Qy	880	SHEHKPIATFEDGSOHLKAGWHAYAAPERPGLAVTSGSQTVFENRLMQGVKGVIPQS	939
Db	658	GLE-----KQDGTIKLV-----DNQNTDNVLTGNGNGTAVTKG	691
Qy	940	GL-TVKLSAOTGGMTGAEGRKVSXKFSERIRAYAFNPTMSTPRPKNAAYATQHWQGRE	998
Db	692	GFETVKT-----GATDADRQKVT-----VKDATANDADKKVATVKQVATAINSA	735
Qy	999	GLKPLIYEMOGALIKOLDAHNVHRNAPQPDLOS-----KLETLDLGHEGAEELNDRKF	1051
Db	736	---ATFVKTENLTTSIDEDNPTDNGKDDALKAGDTLTFKAGKNLKVKRDKGKNIITFLAKN	792
Qy	1052	RDELEQSATRSVTVLQGHQGVKLSNGEINSFEPKSPGKALVQSENVNRSQGL-----	1104
Db	793	LEVKTAKVSDTLITCGN-----TPTGGTTATPKVNITTSADGLNFAKETA	837
Qy	1105	----SKSLQK---AVHATPPSAESKLSQMLGHFVSAGVDMSHQKGEIPL-----	1146
Db	838	DASGKNVYLKGIATTLTEPSAGAK-----SSHVDLNVDAITKKSNAASIEDVLKAGWNIQ	892
Qy	1147	--GQRD-----PNDKTALTKSRLILDVTI-----GELHELADKAKLVSDHK	1187
Db	893	GNGNVDYVATYDTVNFDTDDSTGTT-----TVTVTKADKGADGVKIGAKTSVIKDNH	945
Qy	1188	PDADQIKQLRQOFDTLREKRYESNPVHYTDMGFTHNKALEANYDVAKAFINAPKE---	1244
Db	946	KLFTGKDLKQ-----ANNGATVSEDDG-----KDTGTGLVTAKTVIDAIVKSGWR	991
Qy	1245	--HHGVNLTTRTVLESQGSAAELAKKLNTLLSLDLSGESMSFSRSYGGGVSTVFVPTLSKK	1302
Db	992	VTGEGATAETGATAVNAGNAE-----TVTSGTGVNFK---NGNATITATV-----	1032
Qy	1303	VPVPVTPGAGITLDRAYNLFSFRTSGGLNVSFORDGVGSVINIVATGHDVWP-----	1357
Db	1033	-----SKDNGNINVKY--DVNVGDLKIGDDKKIVADVDTTLTTLVT	1069
Qy	1358	GKKTSGNAGSDWLSAXHKISPDRLRIGAAVSGTLOGLTQNSLKKFLTEDELPGFHLGTHG	1417
Db	1070	GGKVSVPAGANSVNNKKL-----VNAEGLATALNNLSWTAKADK---YADGESEG	1117
Qy	1418	TLTPAELLQKGIHOMKQSGSKLTFSDTSANLDRAGINLNEGSK-----PNGVTAR	1470
Db	1118	E-----TPQEVKAGDKVTF-----KAGKNLKVQSEKDFYISLQDTLTGL	1157

1471 USAGLSANLAAGSRERSTTSGQSGSTTSANNRPTFLNGVAGAGANLTAALGVARSSTH 1530  
1158 TSITLGGTAN---GRNDTGVINKOGLTITLAN-----GAAAGTDSNGNTISVTIK 1205  
1531 EGKPGVGFPAFTSTNVAAL-----ALDNRSTQSISLELRABPV-----TSNDISE 1577  
1206 DGISAG--NKEITNVKSALKTYKDTQNTADETDQKEPFAAVKNAVEFVFGNGATVSA 1262  
1578 LSTLGLKHPK--DSATTKMLAALKELDDAK-----PAQLHILQHPFS 1618  
1263 KTDNNGKHTVTDVABAKVGDGLEKDTGKIKLVNDTGNLNTVDATKGSASVAGEFN 1322  
1619 AKDVVGDREYAVNLK--KLVRQQAADSHMSWELSGASHSTTYNNLSRINDGIVELL 1675  
1323 A--VTTDATTAQGTNANERGVVVK-----GSGNATATETDKKVAIVG----- 1364  
1676 HKHFDAALPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMELKGLRQTEKAIL 1735  
1365 ----DVAKAINDAATFVKVENDDSATID-----DSPTDDGANDALKAXDTL--TLKAGK 1412  
1736 DGKVGREVGVLFDQRRNLRVKSVSVQSVKSGEGFNTPALLLGTSNSAAMSMERNIGTI 1795  
1413 NLKVKRDGKNITFALANDLSVKSATVSDKLS-----LGT-NGNKVNITSDTKGL 1460  
1796 NF---KYGODQNTPRFTLEGGIAQANPOVASALTD 1828  
1461 NFAKDSKTGDAN-----IHLNG-----IASTLTD 1485

RESULT 13

US-09-540-236-3459  
; Sequence 3459, Application US/09540236  
; Patent No. 6673910

; GENERAL INFORMATION:  
; APPLICANT: Gary L. Berton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2005-001

; CURRENT APPLICATION NUMBER: US/09/540.236

; CURRENT FILING DATE: 2000-04-04

; NUMBER OF SEQ ID NOS: 3840

; SEQ ID NO 3459

; LENGTH: 2142

; TYPE: PRT

; ORGANISM: M.catarrhalis

US-09-540-236-3459

Query Match 2.2%; Score 207.5; DB 4; Length 2142;

Best Local Similarity 18.3%; Pred. No. 3e-05;

Matches 365; Conservative 292; Mismatches 763; Indels 573; Gaps 93;

8 TEHKAHVTAHNPVGHGVALQQGSSSSPQAAASLAAEGKRGKMPRIHQPSTAADI 67

530 TEVKSIAADAANGQVDQSFNRLTANNKSKYAATI-----XDLH-----GL 572

68 SAAHQOKSFSRLGCLGTGKTSRPAQOPGTHSK-----GATRLDILLARDGGE 118

573 S---QVLTFA-----GDTGTAKLGLILTKGADTRELTDKNGIV 613

119 QHEAAAPDAARLTRSGVKKRRNMDMAGRPVYKGGSGEDKVPQOKRHOLNPNFGWRQM 178

614 KDNG-----NGGGLQVLAKNLSGLDAV-----NTKLTATGKIQVNNNGNTAQIL 659

179 LSKMAHPASANGDRLQHSPPHPOSH-----HEIK-EBPVGSTSKATTAH--ADRV 227

660 -----GGGLTFTQPNIGGNSGKTYGVGDVGFPTDSSGNTPLAGTITYTKEKI 707

228 EIAQEDD-----DSFQOLHOORLARENPPOPEKLGIVAT--PISARFQPKL-TA 275

708 GFAKQDGLSDSKPYLDKDKLVGEVKITKDGINAGKAITGLPSTLNTTYTAPGVHTA 767

276 VAESVLEGTDTTQSP-----LKPQSMKGGAGVTPLATLDKGLQLAPDNPALN--- 327

Db 1HGSTLSNDOKTTRASIAIADVLNAGFNLEGNAGVDFVSTY-----DTVNFINGNA 817  
QY 328 TLLKOTLGLKXDTQHYLAHASSDSQ--HLLLDN--KHLFDIKSTATSYS-----VLHNS 378  
Db 818 TTKAVTYDNKTSK-VAYDVNVBDGTTIHLGTNGKKNQIGVKTTLTTTSAGKNTPIFN 876  
QY 379 HPEIKGKLAQAGTGSVDGSGKISLSGCTQSHNKTMLS-----QPCEAHRSLL 429  
Db 877 KPGD--DDALVKASDIKDNLNTLAGEIRTAGKTASTALQTFSITDEHGNHFTVGNPYSYD 935  
QY 430 TGIWHPAGAARQGESIRLHDDKIHLPELGVWQSAKDTHSOLSRQADGKLYALKDN 489  
Db 936 TSKTPTVTVPAGENGITTKVKNKV-----KVGIDQTKGLTTPKLVGSSNGKGI-- 987  
QY 490 RTLQNLNKSSEKSLVDKIKSVSYVDORGQVAILTDTTPGRHKMSIMPSLDASPESHLSL 549  
Db 988 -----DSKQGNNTITGLSN-----TLTDTATN-----ATTGHVS-- 1015  
QY 550 HPADAHQGLLHGKSELEAQS--AISHLRLVADSEKLFSAALPKQ--DGNELKMKAM 605  
Db 1016 -----EQGLAQAGANKTRAASIGDVLNAGFNLQNGEAKDFVSTYDTVNFIDGNATTAKVT 1071  
QY 606 PQHALDEHFCHDHOISGFPHDDHGQNALVKNFRQOHACPLGNNDHOFHPGWNLTALVI 665  
Db 1072 -----YDDTKQISTVTYD-----VNVD-----NKTLEVTG 1096  
QY 666 DNQLGLHHTNPEPHEILDMGHLSLALQEGKLHYFDOLT--GWTGAESDCKQL 717  
Db 1097 DKXGLVKTIT-----LTTSANGNATKFSAADGDA--L 1127  
QY 718 KGLDGAAYL-LKDGVEVKRL--NINOSTSSIKH-----GTENVFSLPHVRNKPPEPDAQLGL 771  
Db 1128 VKASDIATHLSTLAGDIQTAKGASQASSASYVDADGNKVIY----- 1169  
QY 772 NKDDKAQAMAVIGVKNKYLALTEKDIRSFQIKPGTQ-----QLERPAQTLSREGISGLK 826  
Db 1170 DSTDK-----KYQVNEKQVD--KITEVTDKLVAQAQTPDGLTAAQMVKSVIN 1217  
QY 827 DIHYDHQNLVALTHEGEVFPHPREAWNGAESSWHKIALPOSESKLKSLDMSHEHKPI 886  
Db 1218 KEQVNDANKKQGINED-----NAFVKGLEKAA-----SDNKTNAAVT-----V 1256  
QY 887 ATPEDGSOHLKAGWHAYAAPERGPLAVGSGSTVFNRLMQGVKGVIPGSGLTVKLS 946  
Db 1257 GDLANAVAQTPLTFFAGDTGTTAKLGE--TLITKGGQDTNKLTDNNIGVAGTGDGTVKLA 1315  
QY 947 AQTGGM--TGAEGRKVSFKSERIRAVAFNPTMTPTPIKNAAYATQHGQREGKLPILY 1004  
Db 1316 KDLTNLSVNAAGTTRIDEKGISFVDA--NGQAKANTFVLSANGLNMG--KRIS 1365  
QY 1005 BMOGALIKOLDAHNVNRHNAPOFDIOSKLETLDLGEHGAELNDMKRPFDELEQSATRSVT 1064  
Db 1366 NI-GAAVDNDNAVNPQ-----FNEVAKTVNNLNNQNSGAS 1401  
QY 1065 ---VLGQHOG--VLKSNGEINSEKPSPK--ALVQSFNVNRSGO---DLKSLSQAAVH 1113  
Db 1402 LPFVVYTDANGXPINGTDGKPOKAIKGADGKYVHANGVFVVDKGNPITDADKANLAH 1461  
QY 1114 ATPPSAESKLSMLGHFVSAGVDMSHQKEIPLGRQDRPNDKTALTATKRLILDTVTIGEL 1173  
Db 1462 KXPLDAGHQVVASLGG--NSDAITLTIKSTLP--QIATPNTG-----NNTNAGOA 1508  
QY 1174 HELADKALIVSDHKDPDQIQOLROQFDTLREKRVESNPVHYTDMGTNKAALANYDA 1233  
Db 1509 QSL---PSLSAAQQAQNAASVKDVLNVGNFNLQTHNHQVDFVKAYDTVNFVNGTG--ADITS 1563  
QY 1234 VKAFINAFKKEHHGNLTATTVL-----ESQGSALAKK-----LKNLTL---SLDGSMS 1282  
Db 1564 VRSADGTMW-----NITVNTALATDDGCVNLIRAKOGKFKADDMENGSLKAGKSAS 1617  
QY 1283 FSRSS-----YGGGVSTVFVFTLSKKV-----PVPVTPGAGITLDRAYNLSF 1323

Db 1618 DAKTPTGLSIVNPADKSTGDAVALNNLSKAVFKSKDGTITTTTSSDGSISQKDNSSI 1677  
 QY 1324 SRTSGGLNVFGRDGVGSGNIMVATGHDVPMYTKTKTSAGNASDWLSAKHKISPLDRIG 1383  
 Db 1678 TLSKDLGVG---GKVISNVGKT-----KDTDAAN-----VQQLNEVRNLLGLG 1719  
 QY 1384 AAVSGTGLQTLNLSKFKLTELPGFIHGLT---HGTL-----TPABELKQIEHOMK 1434  
 Db 1720 NAGNDONADGNQNIADIK--KDPNSGSSNRIVIKAGTVLGGKGNNDTEKLTAG----- 1771  
 QY 1435 QGSKLTFSDVTSANLDEL-RAGINLNEGSKPGVTVARVSAG-LSASANLAAGSRERSTTS 1492  
 Db 1772 -GVQGVGDKDGNAGDLSNVWTKQDGSKKALLATYNAAGQTNVLTNNPASAIDRINQ 1830  
 QY 1493 GQFGSTTSASNNRPTF--LNGVCAGANLTPAALGVVAHSSTHEGKPVGIFPAFTSTNVSAAL 1550  
 Db 1831 GIRFFHVNDGNQEPVVQGRNGIDSSASGKHSVAIGFOAKADGE-----AAV 1876  
 QY 1551 ALDNET---SOSISLELKAEPVTSNDISELTSTL--GKH-----FKDSATTKMLAALKELD 1602  
 Db 1877 ALGROTQAGNSIATG-DNAQATGQDSIAIGTGNVVGKHSAGIADGPPSTVK----- 1926  
 QY 1603 DAKPAQLHILQOHFSAKDVVDGDERYEAVRNLKLVIRQQAADSHSMELGSASHSTTYNN 1662  
 Db 1927 -----ADNSYSVGNV-----NQFTDATQTDVFGVG-----NN 1953  
 QY 1663 LSRINDGIVELHKKHFDAAIPASSAKRLGEMMNDPALKDIIKQLQSTPFSSASVSMEL 1722  
 Db 1954 ITVTESNV-----ALGNSAISAGTHAGTQAKKSDGTAGTTITAGATGVK--- 2000  
 QY 1723 KDLREOTEKALLDKGVREEVGLVFDNRNLRVKSVSQ--SVKSGGFTNPAALLGTS 1781  
 Db 2001 --GFAGQT-----AVGAVSGASGAER---RIGNVAAGEVSATSTDAVNGSGLYKATQ 2048  
 QY 1782 NSAAAMERNIGTINFKYQDQD-----TPRFTLEGGLAQANFQ- 1821  
 Db 2049 SIANTNE---LDHRHONENKANAGISSAMAMASMPQAVIPGRSMVTGGIATHNGQG 2103  
 QY 1822 -VASALTDLKEG 1833  
 Db 2104 AVAVGLSKLSDNG 2116

RESULT 14  
 US-08-968-685A-10  
 ; Sequence 10, Application Us/08968685A  
 ; Patent No. 6214981  
 ; GENERAL INFORMATION:  
 ; APPLICANT: TUCKER, KENNETH  
 ; APPLICANT: PLOSILA, LAURA  
 ; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE  
 ; TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: PENNIE & EDMONDS LLP  
 ; STREET: 1155 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10036-2711  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/968,685A  
 ; FILING DATE: No. 6214981ember 12, 1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Baldwin, Geraldine F.  
 ; REGISTRATION NUMBER: 31,232  
 ; REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 790-9090  
 ; TELEFAX: (212) 869-8864  
 ; TELEX: 66141 PENNIE  
 ; INFORMATION FOR SEQ ID NO: 10:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2123 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; MOLECULE TYPE: peptide  
 ; US-08-968-685A-10

Query Match 2.2%; Score 205.5; DB 3; Length 2123;  
 Best Local Similarity 18.5%; Pred. No. 4.1e-05;  
 Matches 332; Conservative 251; Mismatches 608; Indels 601; Gaps 94;

QY 388 AQAGTGSYSVDKSG-----KISLGS-----GTQS-----HNKTMLSQPCGAHRSLLT 430  
 Db 65 AVAGIGISEADGGKGANARGDKSIAIGDIAQALGSQSIAGDNKIVHNSNNAN----- 119  
 QY 431 GIWQHPAGAAPOGESIRLHDDKIHILPELGVWQS---ADKDTHSQLS-----ROA 479  
 Db 120 -----IGAKASGNESIAIGGDVLAASHASIAIGSDDLKYLKKTQQISELLPIIRGOKA 173  
 QY 480 DGKLYALKDNRTLQNLSDNKSSEKLVDKIKSVSDQRGQVAILTDT-----PGRHMSIMP 535  
 Db 174 LNDIYQLADT-NLQYRTHAGHASTAVGAWSY-AKGHFSNAFGTRATAGTYSVLAVGL 231  
 QY 536 SLDASPESHISL-----SLHFADAHQGLLHGKSELEA-QSVAISHGRLVWADSECKLPSA 589  
 Db 232 TATAKAASSIAVGSNAQAIQFAATAVG---GSTQVNLNRGIALGFG-----SQ 276  
 QY 590 AIPQGDGNELKMKAMPOHALDEHFDHHDQISGFFHDDHQQALNALVKNFRQOQHACPLGN 649  
 Db 277 VLQKNDNVNAANVRA-----YAPDDNPIDNRYKATFN----- 310  
 QY 650 DHQFHPGNWLTALVIDNQLGLHHTNPEPHEILDWHLGS-----LALQEGKHLHFD 701  
 Db 311 -----GATDFVSGNSNG---NDSIRRKLIINVG-AGSADTDAVNVAQKEAVRLANR 358  
 QY 702 QLTGWTGAESDCKQLKKGL-----DGAAYLLKDGVEKRLNINQSTSI 745  
 Db 359 QIT--FKGDDSN-NRVEKGLKTLITGAGTSALTQHNIGVQVQDGLKVLQATLTL 415  
 QY 746 KHGTENVFSLPHVRNKPDPGALQGLKNDKQAQAMAVIGNKYLALTEK---GDIRSFQ 801  
 Db 416 KMT-----TENLTANEK-----VTVGKTRLTDDKIGFTNDMNGID 451  
 QY 802 -----IKPGTOOLER-PAQTLRSREGIS-GEKLDHVDHKQNLIALTHEGEVPHQ 849  
 Db 452 ESKPYLDKDTGIHAGGQKTKLTAGVQVDDDAATYQGLKKVNTQTAESALQTF----- 503  
 QY 849 PREAWQNGAESSSWHKLALPQSESKLKSLDMSHEHKPIAT-----FEDGSQHLKAG 900  
 Db 504 VKKVDKNGDA-----NDSKI---ITVKNKPKDGTQVNTLKLKGENGVDTVTETN 551  
 QY 901 GWHAYAAPERGPLAVGTSG-----SQTVFNRLMQ-GVKGKVI-----PGSGLT-- 942  
 Db 552 GTVTFGLNQNNGLTVGNSTLNDGLSVKNTNSNKQIQVGADGITFTDTSNKPAGAGIENT 611  
 QY 943 -----VKLSAQTGGM-----TG-AEGRKYSSKESERIRAYAFNPT----- 976  
 Db 612 TRITROGIGFANNITGSLDANKPLTPTGTINAGKELTNVQS-----AINPATNGGGLDF 665  
 QY 977 ---MSTPRPIKNAAYATQHGQWQREGKLPLEYMQ-----GALIK--- 1012  
 Db 666 MNRSLTANTEKSGSAAT-----IKDLYNLSQVPLTFAGDTGNTVTKLGEILKVKG 716  
 QY 1013 -----QLDAHNVNRHNAPOPD--LQSKLE-----TLDLGEHGAEL 1044  
 Db 717 GKTTADDLTKNNIGVVADSTONSILTAKLTSLDLAVNTKTLTASDKVTVDSGNNTAKL 776





This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:20:22 ; Search time 88.3207 Seconds

(without alignments)  
6477.980 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAHN.....NPQVASALTDLKXGLEMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	9448	100.0	1838	9	US-09-879-248-8
2	9448	100.0	1838	15	US-10-441-736-8
3	325	3.4	1957	12	US-10-365-742-120
4	275	2.9	2514	12	US-10-282-122A-66121
5	275	2.9	2514	15	US-10-320-800-40
6	273.5	2.9	6713	12	US-10-282-122A-43811
7	267.5	2.8	2659	14	US-10-311-879-28
8	262	2.8	6281	9	US-09-815-242-12996
9	261	2.8	2799	12	US-10-282-122A-65564
10	258.5	2.7	6641	12	US-10-282-122A-70580
11	258.5	2.7	10203	16	US-10-661-809-23
12	257	2.7	5795	9	US-09-815-242-12610
13	256.5	2.7	2398	12	US-10-282-122A-70176
14	253	2.7	2045	12	US-10-282-122A-74463
15	248.5	2.6	3158	9	US-09-815-242-12611

16	247.5	2.6	2344	9	US-09-815-242-12713	Sequence 12713, A
17	245	2.6	2703	12	US-10-282-122A-66108	Sequence 66108, A
18	243.5	2.6	3225	16	US-10-408-765A-254	Sequence 254, App
19	242	2.6	2434	9	US-09-815-242-5835	Sequence 5835, Ap
20	241	2.6	1577	12	US-10-282-122A-69069	Sequence 69069, A
21	240.5	2.5	2086	9	US-09-815-242-5639	Sequence 5639, Ap
22	240.5	2.5	5560	12	US-10-263-929-142	Sequence 142, App
23	239	2.5	2712	12	US-10-282-122A-67070	Sequence 67070, A
24	238	2.5	2283	14	US-10-172-502-4	Sequence 4, Appli
25	237.5	2.5	1905	15	US-10-259-194A-86	Sequence 86, Appli
26	236	2.5	2271	12	US-10-282-122A-43924	Sequence 43924, A
27	236	2.5	3241	9	US-09-841-786-1	Sequence 1, Appli
28	236	2.5	3241	12	US-10-647-057-1	Sequence 1, Appli
29	234.5	2.5	2437	9	US-09-815-242-5834	Sequence 5834, Ap
30	232	2.5	1981	9	US-09-928-457-38	Sequence 38, Appli
31	230.5	2.4	1491	12	US-10-282-122A-65570	Sequence 65570, A
32	229.5	2.4	4273	12	US-10-282-122A-66115	Sequence 66115, A
33	229.5	2.4	4322	16	US-10-437-963-104793	Sequence 104793, A
34	228	2.4	3533	12	US-10-282-122A-70177	Sequence 70177, A
35	228	2.4	3975	16	US-10-437-963-165014	Sequence 165014, A
36	227.5	2.4	2541	14	US-10-177-293-470	Sequence 470, App
37	227	2.4	3286	12	US-10-282-122A-49697	Sequence 49697, A
38	226.5	2.4	2025	9	US-09-815-242-5703	Sequence 5703, Ap
39	225	2.4	3692	12	US-10-282-122A-71235	Sequence 71235, A
40	223.5	2.4	1479	16	US-10-437-963-106122	Sequence 106122, A
41	223.5	2.4	2124	15	US-10-369-493-11841	Sequence 11841, A
42	222	2.3	4823	15	US-10-051-874-169	Sequence 169, App
43	221.5	2.3	5262	15	US-10-051-874-167	Sequence 167, App
44	221	2.3	2265	12	US-10-282-122A-45123	Sequence 45123, A
45	220.5	2.3	4952	15	US-10-051-874-56	Sequence 56, Appli

## ALIGNMENTS

### RESULT 1

US-09-879-248-8  
; Sequence 8, Application US/09879248  
; Patent No. US20020062500A1  
; GENERAL INFORMATION:  
; APPLICANT: Fan, Hao  
; APPLICANT: Wei, Zhong-Min  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE  
; FILE REFERENCE: 21829/81  
; CURRENT APPLICATION NUMBER: US/09/879,248  
; PRIOR FILING DATE: 2001-06-12  
; PRIOR FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 1838  
; TYPE: PRT  
; ORGANISM: Erwinia amylovora  
US-09-879-248-8

Query Match	100.0%;	Score 9448;	DB 9;	Length 1838;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1838;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MELKSLGTEHKAHVHTAHN	PVGHVALQOGSSSSPQNAASLAAGKRGKMPRIHQ	60
Db	1	MELKSLGTEHKAHVHTAHN	PVGHVALQOGSSSSPQNAASLAAGKRGKMPRIHQ	60
QY	61	STAADGISAHQKQKSFSLRGCLGKTKFGRSAPQGGTTHSKGATLRDLLARDDEGTOH	120	
Db	61	STAADGISAHQKQKSFSLRGCLGKTKFGRSAPQGGTTHSKGATLRDLLARDDEGTOH	120	
QY	121	EAAAPDAARLTRSGGKVRKRNMDMAGRPVMYKGGSGEDKVPYTOQKRHLNFFGQMRTMLS	180	
Db	121	EAAAPDAARLTRSGGKVRKRNMDMAGRPVMYKGGSGEDKVPYTOQKRHLNFFGQMRTMLS	180	

181 KMAHPASANAGRLQHSPPHIIQGSHEIKEEPVGSSTKATTAHADRVETIAQDDDDSEFOQ 240  
 181 KMAHPASANAGRLQHSPPHIIQGSHEIKEEPVGSSTKATTAHADRVETIAQDDDDSEFOQ 240  
 241 LHQORLARERENPPQPKLGVAIPISARPOPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300  
 241 LHQORLARERENPPQPKLGVAIPISARPOPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300  
 301 GAGVTPPLAVTLKQKLOLAPDPPALNTLLKOTLGKDTQHYLAHASSDGSQHLLDNKG 360  
 301 GAGVTPPLAVTLKQKLOLAPDPPALNTLLKOTLGKDTQHYLAHASSDGSQHLLDNKG 360  
 361 HLFIDIKSTATSYVLNHSHPGEIKGLAQAGTGSVSDGSKGKISLGSGTQSHNKTMLSQ 420  
 361 HLFIDIKSTATSYVLNHSHPGEIKGLAQAGTGSVSDGSKGKISLGSGTQSHNKTMLSQ 420  
 421 PGSARSLITGIWOHPAGAAPPGESIRLHDDKIHIHLPBLGVWQADKDTSHLSROAD 480  
 421 PGSARSLITGIWOHPAGAAPPGESIRLHDDKIHIHLPBLGVWQADKDTSHLSROAD 480  
 481 GKLYALKDNRTLQNLSDNKSSEKLVDKISYSDVQDQGOVAILTDTGRRHKNSTMPSLDAS 540  
 481 GKLYALKDNRTLQNLSDNKSSEKLVDKISYSDVQDQGOVAILTDTGRRHKNSTMPSLDAS 540  
 541 PESHISLSLHFADAHQGLHKGSELEAQSVASHGRVLVADSEKLPFSAAIPOKQDGNEL 600  
 541 PESHISLSLHFADAHQGLHKGSELEAQSVASHGRVLVADSEKLPFSAAIPOKQDGNEL 600  
 601 KMKAMPOHALDEHFGHDHQTSGFFHDDHGOINLVKNFRQOAHACPLGNDHQHPGWNLT 660  
 601 KMKAMPOHALDEHFGHDHQTSGFFHDDHGOINLVKNFRQOAHACPLGNDHQHPGWNLT 660  
 661 DALVIDNOLGHHNTNPPEHETLDMGHLGSLALQEGKLYPDQLTKGWTGAESDCKQLKKG 720  
 661 DALVIDNOLGHHNTNPPEHETLDMGHLGSLALQEGKLYPDQLTKGWTGAESDCKQLKKG 720  
 721 LDGAAYLLKQGEVRLMINOSTSSI KHGTENVFLPHVRNKPDPGDALQGLNKDDKAQAM 780  
 721 LDGAAYLLKQGEVRLMINOSTSSI KHGTENVFLPHVRNKPDPGDALQGLNKDDKAQAM 780  
 781 AVIGVKNYALATEKGDITRSFOIKGTQOQLERPAOTLSREGISGELKDIHVHDKONLYALT 840  
 781 AVIGVKNYALATEKGDITRSFOIKGTQOQLERPAOTLSREGISGELKDIHVHDKONLYALT 840  
 841 HEGEVFQPREAWQNGAESSEWHKALPOSESKLSLDMSEHKPIATFEDGSOHLKAG 900  
 841 HEGEVFQPREAWQNGAESSEWHKALPOSESKLSLDMSEHKPIATFEDGSOHLKAG 900  
 901 GWHAYAAPERGLAVGTSGSTVFNRLMQGVKGKVIPSGGLTVKLSAQGTGWTGAERKV 960  
 901 GWHAYAAPERGLAVGTSGSTVFNRLMQGVKGKVIPSGGLTVKLSAQGTGWTGAERKV 960  
 961 SSKFSERIRAVAFNPTMTPRPIKNAAYATOHQOEGREGLKPLIYEMOGALIKOLDANVR 1020  
 961 SSKFSERIRAVAFNPTMTPRPIKNAAYATOHQOEGREGLKPLIYEMOGALIKOLDANVR 1020  
 1021 HNAPOPDLOSKLETLDLGEHGAELNDMKRRFDELEQSAATRSVTVLGOHQGVLSKNGEIN 1080  
 1021 HNAPOPDLOSKLETLDLGEHGAELNDMKRRFDELEQSAATRSVTVLGOHQGVLSKNGEIN 1080  
 1081 SEFKPSGKALVQSFNVNRSGDLSKLSQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 1081 SEFKPSGKALVQSFNVNRSGDLSKLSQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 1141 KGEIPLGRQDNDKTALTLSLIDVTITIGELHELADKALVSHKPDADQIKOLROOF 1200  
 1141 KGEIPLGRQDNDKTALTLSLIDVTITIGELHELADKALVSHKPDADQIKOLROOF 1200  
 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260  
 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQ 1260  
 1261 SAELAKKLKNTLLSLDSGSMFSRSYGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320

1261 SAELAKKLKNTLLSLDSGSMFSRSYGGVSTVFVPTLSKKVPVPIPGAGITLDRAYN 1320  
 1321 LSFRTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTKKTSGAGNASDMLSAKHKISPD 1380  
 1321 LSFRTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTKKTSGAGNASDMLSAKHKISPD 1380  
 1381 RIGAAVSGTLOQTLONSLKFKLTEDELPGFHGLTHGTLTPAELLQKGIHQMKQSKLT 1440  
 1381 RIGAAVSGTLOQTLONSLKFKLTEDELPGFHGLTHGTLTPAELLQKGIHQMKQSKLT 1440  
 1441 FSVDTSANLDRAGINLNEDEGSKPNVGTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500  
 1441 FSVDTSANLDRAGINLNEDEGSKPNVGTARVSAGLSASANLAAGSRERSTTSQFGSTTS 1500  
 1501 ASNNRPTFLNGVAGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
 1501 ASNNRPTFLNGVAGAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
 1561 SLELKAEAPVTSNDISELTSTLGHKFDKSAATTWMLAALKELDDAKPAEQHIIQOHFSAK 1620  
 1561 SLELKAEAPVTSNDISELTSTLGHKFDKSAATTWMLAALKELDDAKPAEQHIIQOHFSAK 1620  
 1621 DVVGDERYEAVERNKLKLVIRQQAADSHMELGSAHSTTYNNLSRINNDGIVELLHKHFD 1680  
 1621 DVVGDERYEAVERNKLKLVIRQQAADSHMELGSAHSTTYNNLSRINNDGIVELLHKHFD 1680  
 1681 AALPASSAKELGEMMNNDPALKDIIKQLOSTPFSSASVSMELKDGLREOTEKAILDKGVG 1740  
 1681 AALPASSAKELGEMMNNDPALKDIIKQLOSTPFSSASVSMELKDGLREOTEKAILDKGVG 1740  
 1741 REEVGULFQDRNNLRVKSVSQSVSKSEGFNTPALLLGTNSAAMSVERNIGTINFKYG 1800  
 1741 REEVGULFQDRNNLRVKSVSQSVSKSEGFNTPALLLGTNSAAMSVERNIGTINFKYG 1800  
 1801 QDQNTPRFTLEGIAQANPOVASALTDLKKEGLEMS 1838  
 1801 QDQNTPRFTLEGIAQANPOVASALTDLKKEGLEMS 1838

RESULT 2

US-10-441-736-8  
 ; Sequence 8, Application US/10441736  
 ; Publication No. US20040016029A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wei, Zhong-Min  
 ; APPLICANT: Schading, Richard L.  
 ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS  
 ; TITLE OF INVENTION: RESISTANCE  
 ; FILE REFERENCE: 21829/203 (EBC-003)  
 ; CURRENT APPLICATION NUMBER: US/10/441,736  
 ; CURRENT FILING DATE: 2003-05-20  
 ; PRIOR APPLICATION NUMBER: 60/107,243  
 ; PRIOR FILING DATE: 1998-11-05  
 ; PRIOR APPLICATION NUMBER: 09/431,614  
 ; PRIOR FILING DATE: 1999-11-02  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 8  
 ; LENGTH: 1838  
 ; TYPE: PRT  
 ; ORGANISM: Erwinia amylovora  
 US-10-441-736-8

Query Match 100.0%; Score 9448; DB 15; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MELKSLGTEHKAHVTAHNVPVGHVALQOQSSSSSPQNAASLAAGKNGKMPRIHQ 60  
 DB 1 MELKSLGTEHKAHVTAHNVPVGHVALQOQSSSSSPQNAASLAAGKNGKMPRIHQ 60  
 QY 61 STAADGISAAHQKKSFSIRGLGCLGTFKFSRGAPOQPGTTHSKGATLRLDARDGETQH 120



Db 61 STAAAGISAHQKKSFSURGLCTKFSRSAPQPGTTHSKGATRLDLLARDGETQH 120  
QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGSGEDKVPTQOKRHLNNFQOMQTMLS 180  
Db 121 EAAAPDAARLTRSGGVKRRNMDMAGRPMVKGSGEDKVPTQOKRHLNNFQOMQTMLS 180  
QY 181 KWAHPASANAGDRLOHSPPHIPGSHHEIKKEEPVGSATTAHADRVEIAQEDDSEFQQ 240  
Db 181 KWAHPASANAGDRLOHSPPHIPGSHHEIKKEEPVGSATTAHADRVEIAQEDDSEFQQ 240  
QY 241 LHQQLARERENPPPKLGVATPISARFOPKLTAVAESVLEGTDTTQSPKPKQMLKGS 300  
Db 241 LHQQLARERENPPPKLGVATPISARFOPKLTAVAESVLEGTDTTQSPKPKQMLKGS 300  
QY 301 GAGVTPFLAVTLDKGLQLAPNDPPALNTLLKQTKGDKTOHYLAHASSDGSQHLLLDNKG 360  
Db 301 GAGVTPFLAVTLDKGLQLAPNDPPALNTLLKQTKGDKTOHYLAHASSDGSQHLLLDNKG 360  
QY 361 HLFDTKSTATSYSVLHNSHPGIBKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420  
Db 361 HLFDTKSTATSYSVLHNSHPGIBKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420  
QY 421 PGEAHRSLTGIWHPAGAAAPQGESIRLHDDKIHILHPELGVWQASADKTHSOLROAD 480  
Db 421 PGEAHRSLTGIWHPAGAAAPQGESIRLHDDKIHILHPELGVWQASADKTHSOLROAD 480  
QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQDQGVAILTDPGRHKMSIMPSLDAS 540  
Db 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDVQDQGVAILTDPGRHKMSIMPSLDAS 540  
QY 541 PESHSISLHPADAHQGLLHOKSELEAQSVAISHGRLVVDSEGLFSAALPKQGDGNEL 600  
Db 541 PESHSISLHPADAHQGLLHOKSELEAQSVAISHGRLVVDSEGLFSAALPKQGDGNEL 600  
QY 601 KMKAMPQHALDEHGHCHDHOISGFHDDHGOALNALKVNNFRQHCACPLGNDHFPGWNLT 660  
Db 601 KMKAMPQHALDEHGHCHDHOISGFHDDHGOALNALKVNNFRQHCACPLGNDHFPGWNLT 660  
QY 661 DALVIDNQLGLHHTNPPEHIELDHGLSLALQEGKLYFPQLTGKWTGAESDCKQLKKG 720  
Db 661 DALVIDNQLGLHHTNPPEHIELDHGLSLALQEGKLYFPQLTGKWTGAESDCKQLKKG 720  
QY 721 LDGAAYLLKQGEVRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
Db 721 LDGAAYLLKQGEVRLNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAM 780  
QY 781 AVIGVNYLALTEKGDIRSFOIKPTQOLERPACTLSREGISGELKDIHVHDKQNLALT 840  
Db 781 AVIGVNYLALTEKGDIRSFOIKPTQOLERPACTLSREGISGELKDIHVHDKQNLALT 840  
QY 841 HEGEVFHPQREAWNGAESSESHKALPQSESKLSLSDMSHEHKPIATFFEDGSQHLKAG 900  
Db 841 HEGEVFHPQREAWNGAESSESHKALPQSESKLSLSDMSHEHKPIATFFEDGSQHLKAG 900  
QY 901 GWHAYAAPERGPLAVGTSGSTVNNRLMQGVKVIIPSGGLTVKLSAQTCGTMCAEGRKV 960  
Db 901 GWHAYAAPERGPLAVGTSGSTVNNRLMQGVKVIIPSGGLTVKLSAQTCGTMCAEGRKV 960  
QY 961 SSKFSEIRIRAYAFNPTWSTPRPIKNAAYATOHGQEGELKPLYEMOGALIKQIDAHNVR 1020  
Db 961 SSKFSEIRIRAYAFNPTWSTPRPIKNAAYATOHGQEGELKPLYEMOGALIKQIDAHNVR 1020  
QY 1021 HNAFPQDLQSKLETLDLGEHGAELLDNMDKFRDLEQSATRSVTVLGHOQGLVKSNGEIN 1080  
Db 1021 HNAFPQDLQSKLETLDLGEHGAELLDNMDKFRDLEQSATRSVTVLGHOQGLVKSNGEIN 1080  
QY 1081 SEFKPSGKALVQSFNNRSGQDLKSQQAQVHATPPESAESKLSQMLGHFVSAGVDMSHQ 1140  
Db 1081 SEFKPSGKALVQSFNNRSGQDLKSQQAQVHATPPESAESKLSQMLGHFVSAGVDMSHQ 1140  
QY 1141 KGEIPLGRQDPNDKLTAKSRLLDVTITGELHELADKAKLVSDHDPDADQIKQLRQOF 1200

Db 1141 KGEIPLGRQDPNDKLTAKSRLLDVTITGELHELADKAKLVSDHDPDADQIKQLRQOF 1200  
QY 1201 DTLREKRYESNPVHYTDMGFTHNKALEANYDAVAFINAFKKEHGVNLTTRTVLESQG 1260  
Db 1201 DTLREKRYESNPVHYTDMGFTHNKALEANYDAVAFINAFKKEHGVNLTTRTVLESQG 1260  
QY 1261 SAELAKLKNLTLLSDGESMSFSRSYGGGVSTVFPVPTLSKKVPVPIPGAGITLIDRAYN 1320  
Db 1261 SAELAKLKNLTLLSDGESMSFSRSYGGGVSTVFPVPTLSKKVPVPIPGAGITLIDRAYN 1320  
QY 1321 LSFSTSGGLNVSFORDGGVGNIMVATCHDVMPYMTGKTSAGNASDWLSAKHKISPLD 1380  
Db 1321 LSFSTSGGLNVSFORDGGVGNIMVATCHDVMPYMTGKTSAGNASDWLSAKHKISPLD 1380  
QY 1381 RIGAAVSTGLOQTLONSLKFKLITDELPGFTHGLTHGLTPAELLQKGIHOMQOGSKUT 1440  
Db 1381 RIGAAVSTGLOQTLONSLKFKLITDELPGFTHGLTHGLTPAELLQKGIHOMQOGSKUT 1440  
QY 1441 FSVDTSANLIDLRAGINLINEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500  
Db 1441 FSVDTSANLIDLRAGINLINEDGSKPNGVTARVSAGLSASANLAAGSRERSTTSGQFGSTTS 1500  
QY 1501 ASNNRPTFLNGVGAGANITAAALGVAHSTHGBKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
Db 1501 ASNNRPTFLNGVGAGANITAAALGVAHSTHGBKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
QY 1561 SLELKRAEPTVNSIDISELTSTLKGKFKDSATTKMLAALKEILDDAKPAQLHLOQHFSAK 1620  
Db 1561 SLELKRAEPTVNSIDISELTSTLKGKFKDSATTKMLAALKEILDDAKPAQLHLOQHFSAK 1620  
QY 1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLKHKFD 1680  
Db 1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGSAHSTTYNNLSRINNDGIVELLKHKFD 1680  
QY 1681 AALPASSAKRLGEMWNNNDPALKDIKQLOSTPFSASVSMELKDGLEQTEKAILDGVKG 1740  
Db 1681 AALPASSAKRLGEMWNNNDPALKDIKQLOSTPFSASVSMELKDGLEQTEKAILDGVKG 1740  
QY 1741 REEVGVLFQDRNNLRVKSVSQSVSKSEGTNTALLGTSNSAAMSERNIGTINFKYG 1800  
Db 1741 REEVGVLFQDRNNLRVKSVSQSVSKSEGTNTALLGTSNSAAMSERNIGTINFKYG 1800  
QY 1801 QDQNTPRRFTLEGGIAQANPOVASALTDLKKEGLEMSK 1838  
Db 1801 QDQNTPRRFTLEGGIAQANPOVASALTDLKKEGLEMSK 1838

## RESULT 3

US-10-365-742-120  
; Sequence 120, Application US/10365742  
; Publication No. US20030204868A1  
; GENERAL INFORMATION:  
; APPLICANT: Colimer, Alan  
; APPLICANT: Alfano, James R.  
; APPLICANT: Cartinhour, Samuel W.  
; APPLICANT: Schneider, David J.  
; APPLICANT: Tang, Xiaoyan  
; TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING  
; TITLE OF INVENTION: NUCLEIC ACIDS, AND USE THEREOF  
; FILE REFERENCE: 19603/4112  
; CURRENT APPLICATION NUMBER: US/10/365,742  
; CURRENT FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/356,408  
; PRIOR FILING DATE: 2002-02-12  
; PRIOR APPLICATION NUMBER: 60/380,185  
; PRIOR FILING DATE: 2002-05-10  
; NUMBER OF SEQ ID NOS: 209  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 120  
; LENGTH: 1957  
; TYPE: PRN  
; ORGANISM: Pseudomonas syringae pv. tomato DC3000  
US-10-365-742-120



```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66121
; LENGTH: 2514
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; US-10-282-122A-66121

Query Match          2.9%; Score 275; DB 12; Length 2514;
Best Local Similarity 18.8%; Pred. No. 9.8e-10;
Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY      7 GTEH-KAAVHTAAHNPV-----GHCVALQQSSS-----SSPQNAAS 43
Db      37 GSAHVKSVPFGTHAPVCRNIFSFLLGFSLCIAVGTANIAFADGIIADKAAPKTOAT 96

QY      44 LAAGEKNRGMRIHQPSTAAADGISAAHQOKKSFSLRGCLGTTKFRSAPQGP----- 98
Db      97 ILQTGNG---IPQVNIQTPTSAGVSNQYQAFDVGNGAILNN--SRSNQTQTQLGGWIOG 151

QY      99 -----TTTSHK-----GATRLDRLARDGETQHEAAAAPDAALTR 132
Db      152 NPWLARGEARVVVNQINSSHSQMGVIEVGRRAEVVIANPAGIAVNGGFINASRATL 211

QY      133 SGGVKRRNMDMAGRPVKG-----GSGED-----KVPTQQRKHQNLNFGQMRQTMLS 180
Db      212 TTQCPQVQAGDLGFKRQGVNVIAGHLGDARDTDFTRILSYHSKIDAPVWGQDVRVAG 271

QY      181 KMAHPASANAGRLQHSPPHPIGSHHIEIKEEPVGVSTSKATTAAHADRVEIAQEDDDSEFQ 240
Db      272 QNDVVAATGNA-----HSP-----ILNNAANTSNNTANNTHIPLFAIDTG----- 312

QY      241 LHOORLARENPFPQPKLVGATPISARFQPKLTAVAESVLEGDTTQSPKLPQSMKGS 300
Db      313 -----KLG-----GMVANKITLISTA-----EQAGIRNOGQLFAS 342

QY      301 GAGVTPPLAVTLDKGKLQAFDPNPALNTLLKQTLGDKDQHYLAHAS--SDGSQHLLLDN 358
Db      343 SGN-----VAIDANGRLVNSGTMAAAN-----AKDTDNTABHKVNIRSQ-----VEN 385

```

```

QY      359 KG-----HLFDIKSTATSYS-----VLHNSHPGEIKGKLAQAGTGSVSDCKSGKI 404
Db      386 SGTAVSQGTQIHQSQSIONTUTLLSSGEILHNS-----GSLKNETSGTI-----EAARL 435
QY      405 SLGSGTQSHNKMTLSQPGEAHRSLLTGIWQHPAGAARPGESIRLHDDKIHLHPGLVW 464
Db      436 AIDTDT-LNNQKGLSQTG-----SQKLHI-----458
QY      465 QSADKDTHSQLSROADGKLYALKONRTLQNLSD-----NKSSEKLVDKIKSYSDVQDQ 518
Db      459 -----DAQGKM-----DNRGRMGLODTAPTASNGSSNQTCN---SYNASFHS 498
QY      519 VAILTDTPGRHKMSIMPSLDASPEHSISLSLHFADAHQGLLHGKSELEAQSVASHGRLV 578
Db      499 TTTPTTATGTATVVISNITAPT-----FAD---GTIRTHGALDNSSSIITANGDT 547
QY      579 VADSEKGLFSAAIKQGGDNELKMKAMPQHALDEHFGHDHQSISGFFHDDHQLNALVKN- 637
Db      548 VSAQGG-LNNAG--QIDIHQLNAKG--SAFDNHG--TIISDAVHIQAGSLNNQNGNI 598
QY      638 NFRQOHACPLGNDHQFHPGNWLTDALVID-NOLGLHTNPPEHILDMGHLSLALQBGK 696
Db      599 TTROQ-----LEITDQLDNAHGKLLSABIAVLAVSSLNQNGE 638
QY      697 LHYFDQLTKGWTGAESDCKQLKGLDGAAYLLKDEVKRLNINOSTSIKHGTENVFSLP 756
Db      639 IATNQQL-----IIHQOOSTAVIDNNTINGTIQSGROVAIOAK 675
QY      757 HVRNKPFGDALQGLNK-----DDKACAMAVIGVKNYKALTEKGDIRSFOIKPQTOOLE 810
Db      676 SLSN-----NGTLAADNKLDIALQDDFVVERNIVAGNE-LSLSTRGSLKNSHTLOAGKRIR 730
QY      811 RPAQTLSPRG-----ISGELKDIHVDHKQNLVALTHEGEVHPQPREAMQNGAESMWHKLA 866
Db      731 IKANNLDNAAQGNIQSGGTTDGTQHN-----LTNRGLIDGQ-----767
QY      867 LPOSEKCLKSLDMSH-----EHKPIATFEDGSOHLKAGGWHAYAAEPGRPLAVGTS 918
Db      768 -----QTKIQAGQWNNITGTGRIYGDNIATRLDNQDENGTA--ATAARENLMIGIG-- 819
QY      919 GSOTVFNRLMQGVKGVIPGSLTVKLSAQTGGMTGAERKVKVSKFSERIRAYAFNPMTS 978
Db      820 -----QLNNRENSLIYSGNDMAVGALDINGQATGKAQR-----853
QY      979 TPPIKNAAYATQHWGQGRGLKPLYEMQGALLKQL-----DAHVRNRNAP 1024
Db      854 -----IHNAGATIEAAGKMRGLGVEKHLNTHNEHLKTQLVETGREHIVDYEAFCRHELLREGT 909
QY      1025 QPDL-----QSKLETLDLGHE-----GAELLNDMK 1049
Db      910 QHELGHSVYNDESDHLRTPDGAAHENWHKYDEKVTQKTQVOTQAPAKIISGNDLTIDGK 969
QY      1050 R-FRDELEQSATRSVTVLGQHOG-----VLKNSGEINSEFKPS-PGKALVQSFN 1096
Db      970 EVENTDSQIIAGNLIVQTEKQGLHNEQTFEGEKVFSENCKLHSYWEKHKGR--DSTG 1026
QY      1097 VNRSQDLSKSLQOAVHATPPSAESKLQSMGLHFVSGVDMSHQKQIPIGRORDPDNDKT 1156
Db      1027 HSEQNTYLPPEIITRNISLGSFAYESHKALSHAPSQGTLPQSNQ---ISLPYTSNFT 1083
QY      1157 ALTKSRL-ILDTVTYIGELHELADKAK-----LVSDHKPDA--DQIKQLRQOQDFTLREKR 1207
Db      1084 PLPSSSLYIINFVKNKYLIVETDPRFANYRQWLGSIDYMLDSLKDLPNNLHKRGDGYEOR 1143
QY      1208 YESNPVKHTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQCSAELAKK 1267
Db      1144 LINEQIAELT--GHRRLDGVQNDDEEQFKALMDNGATAARSMNLVSGIAL-----SAEQVAQ 1197
QY      1268 LKNTLLSLSGSGMSFSRSYGGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSFRTS 1327
Db      1198 LTSDIWMLVQKE----VKLPDGGTQTVLVPQVYVRVKNKGDIDGKGAJ-----LSGSNTQ 1247

```



Db 910 QHELCHSVYNDSDHLRTPDGAHWHKVDYKVTQKTQVOTAPAKIISGNLDITDGK 969  
 QY 1050 R-FRDELEQASRSTVTLQHQHOG-----VLKSNGEINSEFKPS-PGKALVQSFN 1096  
 Db 970 EVFNTDSQIIAGNLIQTEKDLHNEQTGFKKVFSENGKLHYSWREKHGR---DSTG 1026  
 QY 1097 VNRSQDLSKSLQAAVHATPPSAESKLSMLGHFVSAGVDMHOKGEIPLGQRDPNDKT 1156  
 Db 1027 HSEQNTLPEETRNISLSGFAYESRKALSHAPSQGTLPQSGN---ISLPTNSFT 1083  
 QY 1157 ALTKSRL-ILDTVTIGELHELADKAK---LVSDHKPDA---DQIKQLRQFDTLREKR 1207  
 Db 1084 PLPSSSLXIINPVNKGVLVETPRFANYRQWLGSDYMLDSLKLPNNLHKRLDGVYER 1143  
 QY 1208 YESNPVHYTDMGFTTHINKALEANYDAVAFINAFKKEHGVNLTRTVLESQCSAEALAK 1267  
 Db 1144 LINEQIAELT--CHRRLDGYQNDDEQKALMDNGATAARSMNLVSGIAL---SAEQVAQ 1197  
 QY 1268 LKNTLLSLDSGSMFSRSGYGVSTVFPVTLSSKKVPVPIGAGITLDRVNLFSRSTS 1327  
 Db 1198 LTSIDLVLVQKE---VKLPDGGTQTLVLPVYVVRKNGDIDGKAL-----LGSNTQ 1247  
 QY 1328 GGLANVSFGRDGVSGNIMVATGHDVMPYMTGK---KTS------GNASDMSLAKHKL 1377  
 Db 1248 INVSGSLKNSGTIAGNALIINTDILNDIGGRHIAQKSAVATATQDINNIGMLSAEQTL 1307  
 QY 1378 PDLRTGAIV-SGTLOGTQNSLKPCLTEDELPG-FIHGLTHGTLT-----PAELLQKGT 1430  
 Db 1308 --LNAGNINSQSTTASSQNTQSSSTYLDKRMAGIVITGKEGVLAQAQKDIINIAGQIS 1365  
 QY 1431 HQMKGSKLTSVDTSANLTL-----RAGINLEDGSKPNGVTVARVAGLSASA----- 1479  
 Db 1366 NCSEGG-QTRLQAGRDINDLTQVTSKHQATHFDADNHVIRGSTNEVSGSIQKGDVTL 1424  
 QY 1480 --NLAAGSRERSTTSGQSTTSASNNRPTFLNGVAGANLTAALGVAHSSTHEKPVGI 1537  
 Db 1425 GNNLNAAAEVSSANGTL--AVSAKND-----IN-ISAGINTT---HVDDASKHGRSGG- 1473  
 QY 1538 PPAETSTNVSAALADNRTSISLELKRAPVTSNDISELTSTL-----GKHF 1587  
 Db 1474 ---GNKLVTDKAOSHHEATQSTFEGQVVLQAGNDANILGNSVNDSTGTOIQAGNHR 1530  
 QY 1588 DSATTKMLAALKELDDAKPAQLHILQHFSAKQVGVDERYEAVNKLKVIROQAADSH 1647  
 Db 1531 IGTT-----QTSQSETHYQTKSGLSMAGIG-----FTI---GSKTN 1565  
 QY 1648 SMELGSASHSTTYNNLSRINNDGIVELIHKHFDALPASSAKRLGEMM-----NNDPAL 1702  
 Db 1566 TOENQSQSNEHTGSTVSLKGDITI-VAGKHYE-----QIGSTVSSPEGNTIYQAQ 1615  
 QY 1703 DIIKQLQSTPPSSASVSMELKDLRBOETEKAIDLKGVGREVG-----LFDORNRLV 1756  
 Db 1616 SIDIQAAHNKLNSNTTQTYEQKGLTVAFSPVTD--LAQAIAVAQSSKQVQSKND-RV 1672  
 QY 1757 KSVSVSOS--VSKSEGNTPALLGTSNSAAMSERNIGTINFKYQDONTPRFT-LEG 1813  
 Db 1673 NAMAANAGWQAYOTGSAONLANGTINAKQVS-----ISITYGQON--RQTTQVQA 1723  
 QY 1814 GIAQANPQVASALTDLKKEGLEMS 1838  
 Db 1724 NQAQASQIQAGKTKTLIATGAEQS 1748

RESULT 6

US-10-282-122A-43811  
 ; Sequence 43811, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Hasebeck, Robert  
 ; APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John  
 APPLICANT: Carz, Grant  
 APPLICANT: Yamamoto, Robert  
 APPLICANT: Forsyth, R.  
 APPLICANT: Xu, H.  
 TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 FILE REFERENCE: ELITRA 034A  
 CURRENT APPLICATION NUMBER: US/10/282,122A  
 CURRENT FILING DATE: 2003-02-20  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/230,335  
 PRIOR FILING DATE: 2000-09-06  
 PRIOR APPLICATION NUMBER: 60/230,347  
 PRIOR FILING DATE: 2000-09-09  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/267,636  
 PRIOR FILING DATE: 2001-02-09  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 78614  
 SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 43811  
 LENGTH: 6713  
 TYPE: PRT  
 ORGANISM: Staphylococcus aureus  
 US-10-282-122A-43811

Query Match 2.9%; Score 273.5; DB 12; Length 6713;  
 Best Local Similarity 17.8%; Pred. No. 6,7e-09;  
 Matches 407; Conservative 322; Mismatches 904; Indels 651; Gaps 95;

QY 11 KAAVHTAHNPVGHVAL-----QQSSSSSPQNAASLAAGKNGKMPRIHQPSA 63  
 Db 556 KAAVENALSQVTNAKAGLNHNLEQAKSNANTTINGLQHLTTAQQDKLKQ-QVQQAQNV 614  
 QY 64 ADGISAHQKKSFLRGCLGTKEFSRSPQGPQGTTHSKGATLRDLILLARDGETOHEAA 123  
 Db 615 A-GVDTV--KSSANTLANGMTLNSI-----QNTATKNGQNYLD--ATERNKINYNA 664  
 QY 124 APDAARITRGGVKRRNDDMAGRPVKGSGDEKVPDQ--QKRHQL---NNFGQMROTM 178  
 Db 665 VDSANGVINA--TSNPNWD-----ANAINQIATQVTTSTKNALDGTNLTQAKQTA 712  
 QY 179 LSKMAHPASANAGRLQHSPPHIPGSHHEIKEEPVGSSTKATTAHADRV-----EIAQ 231  
 Db 713 TNAIDGATNLNKAQK-----DALKAQVTSAQRVANVTISQQTAN 751  
 QY 232 E-----DDSEFOQLHQORLARENPQPPKLGVAFTPIISARFQPKLT--AVA 277  
 Db 752 ELNTAMQLOHGIDDENATKQTKYRDAEQSKTAYDOAAKAILNKOTGNSDKAAV 811  
 QY 278 ESVELEGTDTTQSPKPSQMLKSGAGVTPPLAVTLDKGLQALAPDNPPALNTLLKQTLGKD 337  
 Db 812 DRALQQVTSKDALNGDAKLAEAKAARQNLTLN---HITNAQRTALEGQINQATTVD 867  
 QY 338 TQHYLAHASS-DGSQHLL-----LDNKGHLFDIKSTATSYSLVHNSHPGETKKG 386  
 Db 868 GVNVTNTNANTLDGAMNSLQGAINDKATLRNQNYLDADSKRNAYTQAVTAAEG-ILNK 926  
 QY 387 LAQAGTGSVSDGSKGKISLGSQTSHNKTMLSQPGEAHSL--LTGIWQHPAGARPOG 444

Db 927 QTGGNTSKADVVALNAVTRAKAALNGAENLNAKTSATNTINGLNPNTLOKDNLKHQV 986  
QY 445 E-----SIRLHDDKTHILHPELVWQOS-----ADKDTHSQLSROA 479  
Db 987 EQAQNVGVGVKDKGNTLNTAMGALRTSITQNDNTTKTSQNYLDASDSKNNTYAVNNA 1046  
QY 480 DGKLYALKD-----NRT-----LQNLSDNKSSEKLVDKI-KSYSDVQDG 517  
Db 1047 NGVINATNPMDANINDMANOVNTTKAALNGAQLAQAKTN--ATNTINNAQDLNQK 1104  
QY 518 QVAILLDTDEGRHKMSIMPSLIDASPESHISLSLHFA-DAHQGLLHGKSELEPAQSAVISHGR 576  
Db 1105 KDALKTOVNAQRVS-----DANNVQHTATELANGAMTALKAAIADKERTKAS-----GN 1153  
QY 577 LVVADSE-----GKLFSAAIKQDGNELKVK 603  
Db 1154 YVNADQEKQAYDSKVNTAENIINGTPNATLVNDVNSAASQVNAAKTALNGDN--LR 1210  
QY 604 AMPQHALLDEHFGHDOISGFFHDDHGQNLALVKNFRQOHAC----- 645  
Db 1211 VAKEHA-----NNTIDGL-----AQLNVQKAKLKEQVQSATLLDGVQTVKNSSTLN 1258  
QY 646 -----PLGDHQPHEPGWNLTD-----ALVIDNQLGLHHTNP--EPHE 680  
Db 1259 TAMKGLRDSIANEATIKAGQNYTDASPNRNEYDSAVTAAKAIINOT-----SNPTWEPNT 1314  
QY 681 ILDMHGLSLALQEGKLYHFDQLTKGWTGAESDCKQ-----LKKGLDGAAYL-- 727  
Db 1315 ITOA--TSQVTTKEHALNGAQLAQAKTAKTNLNNITSINNAQKDALTRNIDGATTVAG 1372  
QY 728 LKDGVEKRLINOSTSIIKHGTENVFSLPHVRNKPPEGDALQGLINKDDKAQAMAVIGVK 787  
Db 1373 VNOETAKATELNNAMHSLONG-----INDETQTKQKYLDAPSXKSAVDQAVNA 1423  
QY 788 YLALTEKGDIRSFOIKPCTQOLEPAPOTLSREGISGEIKDIHVDHKQNLVALTHEGEVFFH 847  
Db 1424 AKAILTKASQGVNDKAAVEQALQNVNSTKTALNGDAKLINEAKAAKQTLGLTLH--INN 1480  
QY 848 QPREAWQ-----GAESSWHKL-----ALPQSSKLSLDMSEHKPIAFFEDGSQ 894  
Db 1481 AQRNALDNEIPOATNVGQNTVAKAQAQLDGMAGOLETSIRDKOTTLQSNYQDADAKR 1540  
QY 895 HQLK-----AGMWHAYAAPERGPLAV-----GTSGSQTVFNRLMQVGKVI 936  
Db 1541 TAYQAVNVAATILINKTAGGNTPKADVERAMQAVTQANTALNGIQNL-ERAKQAANTAIT 1599  
QY 937 PGSGLTIVK--LSAQ--TCGMTGABG-----RKVSSK----- 963  
Db 1600 NASDLNTKQKALKAQVTSAGRVSAANGVEHTATELNTAMTALKRAIADKADTKASGNVY 1659  
QY 964 -----FSERIRAYAF-----NPTMSTPRPIKNA--YATQ 991  
Db 1660 NADANKROAYDEKVTAAEBHIVSGTPTPL-TPSDVTNNAATQVTNAQTOLNGHNLVAKQ 1718  
QY 992 HGWQREGKPLIYEMQAGALIKQLDAHNVNRHNAPOFDLQSKLETLDLGEHGAELND-MKR 1050  
Db 1719 NANTAIDGLTSLNGPKAKLXE---QVQATTLPNVQTVRD-----NAQTNTAMKG 1767  
QY 1051 PRDELEQASATRSVTVLQHQGVKLSNGEINSEFPKSPKALVQSENVNRSQDLSKS-LQ 1109  
Db 1768 LRDSIANEATIKAA--GQNYTDASQKQNDYNNNAVTAAKAI--GQTTSPSMIA 1816  
QY 1110 QAVHATPPSAESKLSQMLGHFVAGVDMSHQGEIPLGRDPNDNDKTALTKSRLLDVT 1169  
Db 1817 QEINAKQVTRAKQALNG-----QENLRTAQTNAKOHLN--- 1851  
QY 1170 IGEELHELA-----DKAKLVSDHDPDADQIKQLRQOPD-----TLREKRYESNPVK---H 1215  
Db 1852 --GLSDLTNAQKDAKROIEGATHVNEVTAQANNADALNTAMTNLKNQIQDQNTIKQGVN 1909  
QY 1216 YTDM-----GPTH-----NKALEAN--YDAVKAFINAKFKEHGVNLTTRTVLES 1258

Db 1910 FTDADBAKKNAYTNVTAQEIILNKAQGTAKGQVETALQNVQRAKNELNGQNVAA-- 1966  
QY 1259 QGSABIAKLKNTLLSLDSGESMSPRSYSGGVSTVFVPTLSKKVPVPIPGAGITLDR 1318  
Db 1967 -NAKITAKVALNNTSINNAQAALKSQIEGATTVAGVQVS-----TMASE 2012  
QY 1319 YNLSFRTSGGLN-----VSFGDGGVSGNIMVATGHDVMPYWTGKTSAGNAS 1367  
Db 2013 LNTAMSNLQGINDEAATAKAAQKYTEADRKQOTAYNDVATAAKTLDDKTAGSNDNKVAVE 2072  
QY 1368 DML-----SAKHISPLDRIGAAVSGTLQ-----GTQNSLKFPLEDELPFGTHLTHGT 1418  
Db 2073 QALQVNTAKTALNGDARLINEAKNTAKQQLATWSHLTNAQKANLTSQIERGTTVAGVQGI 2132  
QY 1419 LTPAEALLQGIHEQMKQSGKLTFSVDT-----ANLDRAGIN-----LNEDG 1461  
Db 2133 QANAGTLNQAM-NQLRQSIASKDATKSESDYQDANADLQNAVNDVATNAEGIISATNPE 2191  
QY 1462 SKPNGVTARVS-----AGLSANLAAGSRERSTTSQGFSTTSASNNRPTFLNG-VGA 1514  
Db 2192 MNPDTTINQKASQVNSAKSALNGEKLAA--VKQTAKSIDIGRLTDLNNAQRTAANAEDVQ 2248  
QY 1515 GANLTAALGVVAHSGSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSISL-----ELKRAEP 1569  
Db 2249 APNLAAYTAAKNKATSLNTAMG-----NLKHALAEKDKNTKRSVNYTDADQPKQAYD 2300  
QY 1570 VTSNDISELSTLGLKHFQKOSATTKMLAALKELDDAKPAEQHLILQHFSAKDVVQDER-- 1627  
Db 2301 TAVTQAEAITNANGSNANE--TQVOAALNQLNQAK-----NDLNGDNKVA 2343  
QY 1628 -----YEAVNKLKLVIROQ-----AADSHSMELGSA----- 1654  
Db 2344 QAKETAKRALASYNLNNNAOSTAATSGIDNATTVADVTAQNTANELNTAMGQLONGIND 2403  
QY 1655 -----SHSTTYNNLSRINDGIVE-----LHKHFDPAAL----- 1683  
Db 2404 QNTVKQOVNFTDADQKKDAYTN-AVTNAQGLDKANGQNNMTRKAQVEAALNQVTAKNAL 2462  
QY 1684 -----PASSAK-RLGEMMN-NDPALKDIILQOSTPFPSSASVSMELK-----DGLREQ 1729  
Db 2463 NGDANVRQAQSDAKANLGLTTLNNAQKQDLTSQIEGATTVNGVNSVTKRAQDLIDGAMQR 2522  
QY 1730 TEKAILDGKVGREEVGLFQDRNNLRKVSVSQS-----VSKSEGFN-----TPALLIGT 1780  
Db 2523 LESAIANKQOQTASENTVIDADPTKKTAFDPAITQAESYLAHDKHTNKDKQAVEQAIQSVT 2582  
QY 1781 SNSAAMSERNIGTINFKYGO-----DONTPRFTLEGGAQANPQVASALTDLKKEGL 1834  
Db 2583 STENALNGDANLQCAKTEATQADIDNLQTLNTPQKTLKQQVNA--QRVSGVTDLKNSAT 2640  
QY 1835 EMKS 1838  
Db 2641 SLNN 2644

## RESULT 7

US-10-311-879-28  
; Sequence 28, Application US/10311879  
; Publication No. US20030186275A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; TITLE OF INVENTION: Antigenic Peptides  
; FILE REFERENCE: toxin  
; CURRENT APPLICATION NUMBER: US/10/311,879  
; CURRENT FILING DATE: 2003-03-17  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 28  
; LENGTH: 2659  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-10-311-879-28

Query Match	2.8%; Score 267.5; DB 14; Length 2659;
Best Local Similarity	18.2%; Pred. No. 3.8e-09;
Matches 374; Conservative 298; Mismatches 780; Indels 599; Gaps 86;	
QY	8 TEHKAHVTAHNPVGHGVALQO--GSSSSSPQNAASLAABGKRGKWPRIHQSPSTAAD 65
DB	933 TOYNAMHNAQQOINTAKTEAQVINNERATPQVSDALT-----KVRAAQTAKID 982
QY	66 GISAHAQKPSLRCGLCTKFSRPAQOGPGTTHSKGATRLDLLARDGGETOHEAAAP 125
DB	983 QAKALLQNKEDNSQ--LVTSKNNLQSSVNOVPST--AGMTQOSIDYNNAKREARETEIT 1037
QY	126 DAARLTRSGGVKRRMDD-----MAGRPVMKGSGEDKVPTQOKRHQLNFMQMRQT 177
DB	1038 AAQRVIDNGDATAQOQISDEKHRVDNALTALNAQKHDLTADTHALEQAVQQLNRTG---T 1093
QY	178 MLSKMAHPASANAGR-----LQHSPPHIPGSHHEIKEEPVGSTSKATTA--HADREVELAQ 231
DB	1094 TTGK--KPASITAYNNSIRALQSDLTSAKNSANALIQPIRTVQEVQSALTNNRV--- 1147
QY	232 EDDSEFFQOLHQ-----QLARERENPPQPKLGVATPISARFQPKLTAV----- 276
DB	1148 --NERLTQAINQLVPLADNSALKTAKTKLDBEINKSVTTDGMTQSSIOAYENAKRAGQTE 1205
QY	277 ---AESVLEGTDTTQSPKPO-----SMLKSGAGVTPFLAVTLDRKGLQLAPD--- 321
DB	1206 STNAQNVINNGDATDQOIAAEKTKVEKYNLSLQAIAGLTPDLAPLQTAKTQLQNDIDQP 1265
QY	322 -----NPPALNTLLK--QTLGKDQTHYLAH-----ASSDGSQH 353
DB	1266 TSTTGMTSASIAFNEKLSAARTKIQEIDRVLASHPDVATIRQNVTAANAAKSALDQARN 1325
QY	354 LLLDNKGHLFDIKS-----TATSYSVLHNSHPGEIKGLAQAGTGSVSDGSKISL 406
DB	1326 GLTVDKAPLENAKQLOYSIDTQSTTGTQDSINAYNAKLTAAENKIQQIN---QVLA 1381
QY	407 GSQTQSHNKMLSQPGEAHRSLLTGIWHPAGAAPQGESIRLHDDKTHIHLPELGVWQS 466
DB	1382 GSPTVEQINTNTSTANQAKSDL-----DHARQALTP-----DKAPLQTAKTQLEQS 1427
QY	467 ADKDT-----HSQLSROADGKL---YALKDNRTLQNLSD-----NKSSEKLV 505
DB	1428 INQFTDTTGMTTASLNAYNQKLAARQKLTEINQVNGNPVTQINQNKVTEANQAKDQLN 1487
QY	506 DKISYSDVQRGQVAIITDTPGRHKMSIMPSLDASPESHISLSLHPADAHQGLLHGKSEL 565
DB	1488 TARQGLTLDRQPALTTL-----HGAS--NLNQAQQNNFTQOINAAQNAHALETIKSI 1538
QY	566 EAQSVAISHRLVVDSEGKLFSAIIPKQGDGNELKMKAMPQHALDEHFGHDHIOISGFFH 625
DB	1539 TALNTAMTKLKDSVAD----- 1554
QY	626 DDHQNLNALVKNFRQOCHACPLGNDHOPFGWNLTDAVIDNQLGLHHTNPEPHEILDMG 685
DB	1555 -----NNTIKSDQNYTDATP--ANKQAYDNVNAAKGVIGET-----TNDVN 1597
QY	686 HLGSLALQEGKLHYFDLTKGWTGAESDCKLKKGLDGAAYLLKDGVEKRLNINQSTSSI 745
DB	1598 TVNKA-----ASVSKTKDALDQOQNLQRAK-----TEATNAI 1630
QY	746 KHGTENVFSLPHVNKPEPGDALQGLNKDDKQAQAMAVIGVKNKLYALTEKGDIFRSFOIKPG 805
DB	1631 THASD-----LNQAQKNALTOQVNSAQNVQA-----VN-----DIKQT 1663
QY	806 TOQLERPAQTLSREGISGELKD IHVD-----HKQNLV--ALTHEGEVHPQPREAWQNG 856
DB	1664 TQSL--NTAMTGLKRGVANHHQVQSDNYVNAATNKKNDYNNAYNHANDII-----NG 1714
QY	857 ABSSSWHKLALPOSESKLSLDMSEHKPIATFEDGSHQOLKAGGWHAYAA----- 907
DB	1715 ---NAQHPVITPDSVNNALSNVTSKEHAL-----NGEAKLNAAKQEAANTALGHLNLMN 1765
QY	908 PERGPLAVGTSGSQTV--FNRLMQGVKGVIPGSGGLTVKLSAQTGMTGAEGRK----- 959

Db	1766	AQRQLSQINGAHQIDAUNTIKQN-----ATNLNSAMGNLRQAVADKQVKRT 1814
QY	960	-----VSSKFESEIRIRAYAFNPTMTSTPRPIKNAAYATOHQHGQREGKLP 1002
DB	1815	EDYADADTAKONAYNSAVSS--AETIINQTTNPTMSVD--DVNRATSAVTSNKNALNGYBK 1871
QY	1003	LYEMOGALIKOLIDAHNVHRNAPOPDLOSKLETLD-----LGEHGAEL-----L 1045
DB	1872	LAQSKTDAARALDALPHLNNAKADVKSKINAASNIAGVNTVKQOQGTDLNTAMGNLQGA 1931
QY	1046	ND-----MKRFRDELESATRSVTVLGHQGVLL--KSNGEINSEFPKSPGKALVQSFNN 1098
DB	1932	NDEQTTLNSQNVQDATPSKKTAYTNAVQAADILKNSGNQNTKQOVTEAMQVNSAKNN 1991
QY	1099	RSQDLSKSLQOAVHATPPSAESKLSQMLGHFVSAGVDMSHOKGEIPIQRQRPNDKTAL 1158
DB	1992	LDGTRLLDQAKQ-----TAKQQLNNMT--HLTTA-----QKTNL 2023
QY	1159	TKSRLLIDVTIGELHELADKAKLVSDHKPDADQIKQLRQQFDTLREKRYESNPVKHYTD 1218
DB	2024	TNQ--INSQTTVAGVQTVQOSNA-----NTLDQAMNTLRQSIANKDATKASED 2068
QY	1219	MGFTHNKALEANYDAVKAFINAFKKEHGVNLTRTVLESQGSAAELAKKLXNTLLSLDSG 1278
DB	2069	Y-----VDANNDKQATAYNNAVAAAEIINANSNPMNPSTITQKAEQVNSKLTALNGD 2121
QY	1279	BMSFSSRSYGGGVSTVPTLSKKVPVPIPGAGITLDRAYNL-----SFRSTSGSLNVS 1333
DB	2122	ENLAAAKQN-----AKTYLNTLT-----SITDAQKNLISQITSATRVSGVDTVK 2166
QY	1334	-----FGRDGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPLDLRG 1383
DB	2167	QNAOHLQOAMASLONGINNESQVSKSEKRYRADTNKQOEDYNAI--TAAKALN----- 2218
QY	1384	AAVSGTLQGLTQNSLKFKL-----TEDELPG-----FIHGLTHGTITLPAELL 1425
DB	2219	---KSTGPNTAQNAVEAALQRVNNAKDALNGDAKLIAAQNAKQHLGLTLH--ITTAQ-- 2271
QY	1426	QKGIEHQMKQSKL--TFSVDTSAN-LD-----LRAGIN-----LNEGSKPNG 1466
DB	2272	RNDLTNQISQATNLGAVESVKQANSLDGAQNLQTAINDKSGTLASQNFLLDADEQKNA 2331
QY	1467	VTARVSAGLSANLAAGSRERSTTSQFGSTTSASNNRPTFLNGVAGAGANLTAALGVAH 1526
DB	2332	YNQAVSAA--ETILNKQTPGNTAKTAVEQ---ALMNVNNAKHALNGT---QNLNNAKQAA-- 2383
QY	1527	SSTHEKPGVGIFFAFTSTNVSAALALDNRTSQSISLELKRAEPVTS-----NDISELST 1581
DB	2384	-----ITAINGASDLNQKQKDALKAQANGAQRVSNQADQVQHNATELNTA 2427
QY	1582	LG--KH-----FKDSATTQMLAALKELDDAK-----PAEQHLILQ 1614
DB	2428	MGLTKHAIADKNTLASSKYVNADSTKQNAVTTKVTNAEHIISGTPVTVPSEVTAAN 2487
QY	1615	QHFSAK--DVVGDERY--EAVRNL-----KKLVIRQQAADSHSMELGSAHSTT 1659
DB	2488	QVNSAKQGEINGDERLUREAKQNTAIDALTQNLTPQKAKUKEQVQANRLE-----DVQTV 2543
QY	1660	YNNLSRINNDGIVELLHKHFDAALPASSAKRLGEMMNDPALDKIIKQLOSTPFSSASVS 1719
DB	2544	QTNQOALNN-----AMKGLRDSIANETTVKTSQNTYTDASPNNQSTYN 2585
QY	1720	MELKD--GLREQTEKAILDGKVBREVGVLFPQDRNNLR--VKSVSVSQSVSKSGEFTNPAL 1776
DB	2586	SAVSNAKGIINQNTNPTMDTSAITQATTVQVNNAKNGLNGAENLRNAQNTAK--QNLNTLSH 2644
QY	1777	LLGTSNSAAMS 1787
DB	2645	LTNNQKSAISS 2655

RESULT 8







Db 1044 VRDYHIETKRIENRPAHITVGNLTASQGNLWLNKDSRI-----VVGGRIITDLDNQ 1097  
 Qy 582 -----SEGLFSAIPKQGD-----GNELKWKAMPQALDHFCHDHOISGEFFH 625  
 Db 1098 KEITNOSTTGKGTDAVGTQWDSVTKKGYSGRRKORREKNT-----SYHDTQL--FTH 1151  
 Qy 626 DHGQNLNVLN-----NFRQOAHAC-----PLGNDHQFHPG--WNLTDALVIDNQLG 670  
 Db 1152 DFTPTPSVIOQNAASPSQPAASAALKIDGASTAAYNGQRIHTGNVSLNNAATVTLNPS 1211  
 Qy 671 LHNTNPEPHEILDMHGLSIALQEGKHYDQITKGTWGAESCKQKGLDGAAYLLKD 730  
 Db 1212 LYTHPD-----NKGW-----LIETDPQFADY--RRWLGSYDMLQQLKLTNHLHRLGD 1259  
 Qy 731 GEVKRLNINOSTSSIKHGTENVFSLPHVRNKPBGDALQCLNKDDRAQAMAVGVNKKLA 790  
 Db 1260 GYTEQKLVN-----EQIHQITGYR-----LDGYKDEE-----QFKA 1292  
 Qy 791 LTEKG--DIRSFOIKPG--TOQ-----LERPAQTLREGISGELKDHYVDHKQ 834  
 Db 1293 LMDNGLTAAKTFTGLTGLIALSAEQVARLTSDIVMENQTVTL-----DGSQTQVLV--P 1345  
 Qy 835 NLVALTHEGVFQPREAWONGAESSWHKLALPOSESKLSLSDMSHEHKPIATFEDGSG 894  
 Db 1346 KVALARKGDL-----NTSGGLISAEQVLLKQNGNLNS--GTIA-----GRQ 1387  
 Qy 895 HOL-----KAGWHAYAAPERGFLAVGTSGSQTVEFNLMOGVKGVIPGSGLTVKLSA 947  
 Db 1388 AVLQARNINSNGNIQADQILGKAEKINSIDGGVOAGRL-----TAQAQNLNLTGTT 1441  
 Qy 948 QTGG-----MTGAEGRKVSSEKSERIRA-YAFNPTWTSTPRPKNAAY-- 988  
 Db 1442 QTSNERNNGTATDRMAGINVGVSHTQVDNRTSDGILSLHAGNDINLNAATVSNQVKG 1501  
 Qy 989 ATOHGWGREGKPLPYEMOQALIKOLDAHNVHNAPODLOSKLETLDLGEHGAELL--N 1046  
 Db 1502 TQITAGNNILNLTITHEAREAYGLDDENHRHVQSTEVGSSIRT-----QNGAMLRAGN 1557  
 Qy 1047 DMKFRDELOSATRSVTVLGOHGVL-----KSNGEINSEFKPSP----- 1087  
 Db 1558 DLKTRQELADEGKTVLAGROVNSEGRQITELDASVSKSGILSSYKTHDPYFSSH 1617  
 Qy 1088 -----GKALV-----QSFNVN-----RSGDLSKSLQOAVHATPPSESK 1122  
 Db 1618 DEAVSGNIGGKMIVAAGQDINVRGSLNLSKGTVLKAGNDIDISTAHNRYTGNEXYESK 1677  
 Qy 1123 LQSMGLHFVSAGVDMHQKEIPLGRQR--DPNDXTALTKSRLT-----DVTIT----- 1170  
 Db 1678 KSGVMG--TGGLGFT-----IGNRKTDDTDRTNIVHTGSIIGSLNGDVTTVAGNRY 1727  
 Qy 1171 ---GELHELADKAKLVSDHPDADQIKLRQOFTLREKYESNPVKHYTDMGFT---HN 1224  
 Db 1728 RQTGSTVSSPEGRNTVTAKSIDVES-----ANNRYATDYVHTREKQGLTVALNV 1776  
 Qy 1225 KALEANYDAVKAFINAFKEHGVNLTTRTVLESQGSAAELAKLNTLLSLDSGESMSFS 1284  
 Db 1777 PVVQAAQNFVQAAQNVGSKNKRNVNMAAANAQW-QYAAQQAQMOQFAPSSSAGQGN 1835  
 Qy 1285 RSYGGGVSTVFVPTLSK-----KVPVPVPGAGITLDRAYNLSFSRTSGG---LNV 1332  
 Db 1836 QSSGISVSVTYGEOKSRNEOKSRVTEAAASQIIGKQT-----TLAATSGSGQSNINI 1888  
 Qy 1333 SFRDGGVSGNIMVATGDVNM--PYMTGKTSAGNASDW-LSAKHKISPDLRIGAAVSGT 1389  
 Db 1889 TGSVDIGHAGTIALDNRHLRQSAQDQSGSEKSKSGWAGVAVKINGIRFGITAGN 1948  
 Qy 1390 L-QOTLQNSLKFTELDPFGIHLTH-----GTLPAELLQKIEHQM 1433  
 Db 1949 IGKKEQGG-----NTTHRHVHTAGQTTIRSGDITTLKGAQLIGIQAQDT 1997  
 Qy 1434 K-----QSKLTFSDVTSANLDRAGINLNEDGSKENGVTARVSAGLSASANLAAGS- 1485  
 Db 1998 RNLHIESVDPTQSKQOQNGNVQTVVGYGFSAGSYRQSKVKADHASVTCQSGIYAGED 2057

Qy 1486 -----RERSTTSGQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHS-----STHEGKPVG 1536  
 Db 2058 GYQIKVRDNTDLKGGIITSSQ-----SAEDKGNLFQATATLTHSDIQNHSRYEKGKSGF 2110  
 Qy 1537 IPPAFSTNVSAALADNRTSQSISLELKRAEPV-----TSNDISELT-----STLGKHPK 1587  
 Db 2111 IGSFPLNG-----GWDGTVTDKQGRPTDRISPAAGYSGDGSKNSTTRSGINTRNIIHT 2165  
 Qy 1588 DSA--TTKMLAALKELD-----DAKPAEQHLI--LQOHFSAKDVVGDREYEAARNLKK 1636  
 Db 2166 DEAGQLARTGRTAKEAREARYTICIDTETADQ--HTGRKNSFD-KDAVAKE-----INLQR 2218  
 Qy 1637 LVIRQ-----QAADSHSMELGSAHSSTYNNLSRINNDGIVLELLHKKHFDPAALPASSAKR 1690  
 Db 2219 EVTKFGRNAAQAAVAADKLGNTQSVRYEQEARTLLEDELOQNTDSEAKAAIRAS---- 2274  
 Qy 1691 LGEM 1694  
 Db 2275 LGQV 2278

RESULT 10

US-10-282-122A-70580  
 ; Sequence 70580, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 70580  
 ; LENGTH: 6641  
 ; TYPE: PRN  
 ; ORGANISM: Staphylococcus epidermidis  
 ; FEATURE:  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (6636)..(6636)

OTHER INFORMATION: X-any amino acid  
US-10-282-122A-70580

Query Match	2.7%;	Score 258.5;	DB 12;	Length 6641;
Best Local Similarity	17.7%;	Pred. No. 8.2e-08;		
Matches 396;	Conservative 319;	Mismatches 805;	Indels 723;	Gaps 90;
QY 33	SSSSPQNAASL-AAGKQNGKMPRIHQPSTAADGISAAHQKKSFLRGLGKTKFFSR	91		
DB 2694	ATQSVQAEALHGAELNQDK-----QTSSTELDGLTDLTDAQRE-KLREQINTNSR	2748		
QY 92	APQ--GQPTTHSKGATLRLDARDG-----ETCHEAADAAR-----	129		
DB 2749	IKQTEQAKALNDAMKKLEQVAQKDGVAHNSDYTNEDSAQDAYNALKQAEIDI	2808		
QY 130	-----LTRSGGVKRRNMDMAGRPVVKGGSGEDKVPQTKRHLNLFQGMROTMLSK	181		
DB 2809	NPFLNAQDITNALNLIKQADNLHGAQKLO-----QDKNTTNAIGNLHNLNPOKDALIQ	2864		
QY 182	MAHPASA-----NAGRLQHPSPHIPGSHHEIK-----	209		
DB 2865	AINGATSRDQVAEKLKEAEALDEAMKQLEDQVNDQDQISNSPFFINEDSDKQTYNDKIQ	2924		
QY 210	--EEPVGSTSKAT-----TAHADRVETIAQEDDDSEFQQLHQORLARE	249		
DB 2925	AKEIINOTSNPTLDKQKIADTLQNIKDAVNNLHGDQ-KLAQSKQDANNQNLNHLDDLTEE	2983		
QY 250	RENPPQPKLGVAATPISARFQPKLTAVAESVLGDTTQSPKQPMKGGSGAGVTPLAV	309		
DB 2984	QKN-----HFXP-----LNNADTRDEVN-----	3003		
QY 310	TLDKQKLOAPDNPALNTLILKQTKGTQHYLAHHSAGSGQHLNLDNKGHLFDIKSTA	369		
DB 3004	-----QLEIAKQLNGDMSTLHKVINDKQIOHLSNVINADNDKKQYDN-----AIK	3050		
QY 370	TSYSLVHNSHPGEIKGLAQAGTGSVSDGKSGKISLGGTSGHNTKMLSPQGEAHRSL	429		
DB 3051	EAEELIHN-HPTDLHKAQ-----DLNKTIDQAHNEL-3082			
QY 430	TGIWQHPAGAAR-----PQGESIR-----LHDDKTHILHPELVGMQS	466		
DB 3083	-----NGESRFKQALDNALNDIDSLNSLVPOQTQVKDNINHVTTLESQAEL-----	3130		
QY 467	ADKUTHSOLSRQADQKLVAKD-----NRTLQNLSDNKSSEKLVDKIKSY-----	511		
DB 3131	-----QKAKELNDAMKMRDSIMNQEQIRKNSNYTINEDLAQQVAYNHAVDKINNII	3183		
QY 512	-----SVDRQGOVAILTDTFGRHKMSIMPSLD	538		
DB 3184	DNATWDPOIILKQATQDINTAINGLKGDKQKQDQAKQITNFTGLTEPQKQAL-----	3238		
QY 539	ASPESHISLSLHFADAHQGLLHGK-----SELEAQSVAISHGRILVAD-----	587		
DB 3239	-----ENILNQTSRANVAKQLSHAKFLNGKMBELKAVAKASLVQRNSNYINEDVSEKEAY	3295		
QY 588	SAALPKQGDGNE-LMKKAMPQHALDERHGHQHSIGFFHDDHQLNALVKNFRQOHACP	646		
DB 3296	EQAIKAK-----GQEIIINSENNPTISSTDIRTTQIEINDAEONLHGD-----	3345		
QY 647	LGNHQHFGHGNLTDALVIDNQLGLHHTNPPEHLLDMCHLGSALQBGKLUHYFDQTKG	706		
DB 3346	AKNEIQNLGINSQAQITKLIQDGRITTTKPAVTKLEBAKAINQAMQOLKOSIADK-----	3401		
QY 707	WTGAESDCKQLKKGIDGAAYLLKQGEVKNLN-----QSTSSIKHGTENVFSLPHVRN	760		
DB 3402	-----DATLNSNYSINESEKKLAYDNVSAQELINQLNDPTMDISINQAITQ	3450		
QY 761	KP-BFGDALQGLNKDDKAQAMAVIGVKNYKALTEKGD-----IRSFQKPGTQOLERP	813		
DB 3451	KVIQAKDSLHCAKLAQCAQADSNLIINGSTNLNDKQKQALNDLNHAGTKQQAIAITAQA	3510		
QY 814	QTLSEGISGLKDIHVD-----HKQNLVALTHEGEVHPQREAWQNGAESSSW-----	862		

DB 3511	NKLINNE--MGTLKTLVEEQSNVHQSKY-INEDPQVQNTYNDISQKREILNGTITDDVLN	3567
QY 863	-HKLALPQSESKLSIDMSHEHKPTATPEDGSGH-----QLKAGWHAYAAPERG	911
DB 3568	NNKIADATQNIHLTKNDLHGDKQKQKQADATNELNLYNLNNSQSQSHDEINSAPSR	3627
QY 912	PLAVTSGTSQTVFNRMLQGVKGVIPGSGLTIVKLSAQTCGTGAGRKVSSKFSERIRAY	971
DB 3628	EVSNDLNHAKAL-NEAMROLENEVALENSVK-KLSDFIN-----EDEAAQNEYSNALQKA	3680
QY 972	--AFNPTMST-----PRPIKNAAYATQGMQGRGLKPLXYEMOGALIKQLD--AHN-----	1018
DB 3681	KDIINGVFSSTLDKATIEDALLELOQVARESLHGEQKLEAKNQVAAEIDNLQALNPGQVL	3740
QY 1019	-----VRHNAPODQLQSKLETLDLGEHGAELNDMKRFRDELEQSATRSVTVLGHQGV	1073
DB 3741	AKTLVNGQASTPEVQEA-----QKAKELNEAMKALKTEINKK-----EQIK	3783
QY 1074	KNGEINSEFEKPSGKALVQSFN--VNRSGQ-----DLSKSLQQAQVHATPPSAESKL	1123
DB 3784	ADSRVYNAD-----SGLQANYSALNYGSOIITATQPELNDKVINRATQITIKTAENNL	3837
QY 1124	--QSMGLHFVSAQ-VDMSHQKGEIPIGRQRPNDKTALTLSRLILDTVTVIGELHELADKA	1180
DB 3838	NGQSKLAELAKSDGNGQSIIEHLQ-----LTQSK-----DKQHDLLINQA	3875
QY 1181	KLVSDBKPDADQI-----KOLRQOFTDLREKRVESNPVKHYTD-----MGTHNKALBA	1229
DB 3876	QT-----KQOVDIIVNNSKQDLSNMNLOQIIVNNDNTVKNQSDFINEDSSQDAYNHAIOA	3931
QY 1230	NYDAVKAFINAFKKEH-----HGVNLTTR-----	1253
DB 3932	AKDLITAHPTIMDKNQIDQAIENIKQALNDLHGSNKLSEKKEASEQQLNLSLNGQD	3991
QY 1254	TVLESQGGAEI-----AKKLNTLLSI-----DSGESMSPSRSYGGGVSTVFVPT	1298
DB 3992	TILNHFISAPTRSQVGEKIASAKQLNNTMKALRDSIADNNETLQSSKYFNEDEQ-----	4046
QY 1299	LSKKVPVPIPCAGITLDRAYNLSPRTSGGLNVSFGDGGVSGNIMVATGHDVWPMYTG	1358
DB 4047	-----QNAVQAVNKAIIIN-----DQPTPMANDEIOSVLNE	4080
QY 1359	KKTSAGN-----ASDWLSAKHKISPDLRGAASVGTQGTQSLQSLK-----	1399
DB 4081	VKQTKDNLHGDKQKLANDKTDAQATLNALYNLQAGRNLETKVQNSNRPPEVQKVQVLAN	4140
QY 1400	-----FKTEDELPGFHLTHGLTTPAELLQKQIEHOMQKQSKLTFSDVTS--ANLD--	1450
DB 4141	QLNDAMKKLDDALTC-----NDAIKQTSNY-INEDTSQQVNFDEY	4179
QY 1451	LRAGINLNEGSKPQNGVTARYSA-----GLSASANLAASRSPRSTTSQGFSGTTS	1500
DB 4180	TDRGNIVAEQTNPMNSPTNITNITADKITEAKNDLHGQKQKQAOOSINTINQMTGLNQ	4239
QY 1501	ASN-----NRPTFLNGV--GAGANLTAALGVVAHSSTHEGKPVGIF	1538
DB 4240	AQEQALQEIQOQOTQTRSEVHOVINKAQALNDSMNTLQSIITDEHEVKQTSVINEVNGQ	4299
QY 1539	PAFTSTVNSAALNDNTSOSI--SLEKRABPVTSN-DISELTSTLGHKFKDSATTKML	1595
DB 4300	TAYNNA-VDRVKQIINQTSNPTNPLEVERA---TSNVKISKDALHGERELNDNKNSTF	4355
QY 1596	AA--LKELDDAKPAEQLHILQOHP-----SAKQVVG-----	1624
DB 4356	AVNHLNDLNAQKEALTHEEQATIVSQVNNIYNKAKALANDMKKLDIVAQQDNVRQSN	4415
QY 1625	-----DERVEAVRNLLKLVIRQQAAD-SHSM	1649
DB 4416	NYINEDSTPQMYNDTINHAQSIIDQVANTPMSHDEIENAINNKHAI---NALDGEHL	4472
QY 1650	ELGSASHSTTYNNLSRIN--NDGIVELLHKHFDAAALPASSAKRLGEMMNDPALKDIK	1706
DB 4473	QAKENANLLINSINDLNAPQDAINELVNE-----AQTRKVAEQLOSAQALNDAMK	4525

```

QY 1707 QLOSTPSSASVSMELK-----DGLREOTEKAILDGKVGREYGVLFQ-----DRNLRV 1756
Db 4526 HLRNSIQNSVQESKYINASDAKQYNHAV-----REVENINEQHPTLDELKIQ 4579
QY 1757 KSVSVQSVSKSGFNTPALLLGTSNAAMS-----MERNIGTFNFKYQDQNTPRPTLE 1812
Db 4580 LTDGVNQA---NNDLNGVELLDADKQNAHQSIPTLMHLNAQAQNALNEKINNAVTRTEVA 4636
QY 1813 GGIAQANPQVASALTDLKKEGLE 1835
Db 4637 AIIQOAK-----LLDHAMENLE 4653

RESULT 11
US-10-661-809-23
; Sequence 23, Application US/10661809
; Publication No. US20040101919A1
; GENERAL INFORMATION:
; APPLICANT: HOOK, Magnus
; TITLE OF INVENTION: BIOINFORMATIC METHOD FOR IDENTIFYING SURFACE-ANCHORED PROTEINS FR
; FILE REFERENCE: P07741US01/EAS
; CURRENT APPLICATION NUMBER: US/10/661,809
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: 60/410303
; PRIOR FILING DATE: 2002-09-13
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 23
; LENGTH: 10203
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-661-809-23

Query Match 2.7%; Score 258.5; DB 16; Length 10203;
Best Local Similarity 17.7%; Pred. No. 1.7e-07;
Matches 396; Conservative 319; Mismatches 805; Indels 723; Gaps 90;

QY 33 SSSPQNAASL--AAGKGRCKPRIHPSTAAAGISAAHQOKSFSLRGCLGPKTKFSRS 91
Db 6262 ATQSVQNAQALHGAELNODK-----QTSSTELDGLTDLTDAQRE-KLREQINTNSRDD 6316
QY 92 APQ--GPGTTHSGATLRDLARDG-----ETOHEAAPDAAR----- 129
Db 6317 IKQIEQAKALNDAMKLEQVAKQDGVHANSDYTNEDSAQDAYNNAKQAEDINNSS 6376
QY 130 -----LTRSGGVKRNMDMAGRPMVKGSGGKEDVPTQCKRHQLNNFQGMQKTLMSK 181
Db 6377 NPENLAQDITNALNNIKQAQNLHGAQKIQ-----QDKNTTNAICNLNHLNQPOKDALIQ 6432
QY 182 MAHPASA-----NAGDRLQHSPPHIPGSHHEIK----- 209
Db 6433 AINGATSRDQVAEKKEAEALDEAMKQLEDQVODQDINSPPFINEDSDKQTYNDKIQ 6492
QY 210 --EPPVSGTSKAT-----TAHADRVETIAQEDDDSEFOQLHQORLARE 249
Db 6493 AAKEIINTSNPTLDKQKIADTLQNIKDAVNNLHGQ--KLAQSKQDANNQNLHLLDLE 6551
QY 250 RENPPQPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPKPKQSMKLGSGAGVTPAV 309
Db 6552 QKN-----HFKEP-----LINNADTRDYNK----- 6571
QY 310 TLDGKQLQAPNPPALNTLLKQTKDTQHYLAHASSDGSQHLNLLDKGHLFDIKSTA 369
Db 6572 -----QLEIAQQLNGDMSTLHKVINDKQIQIQLHSNYINADNKKQYDN-----AIK 6618
QY 370 TSYVLHNSHPGEIKGKLAQAGTGSVDGKSGKISLGSQTSHNKNKMLSPGFAHRSIL 429
Db 6619 EAEDLIHN-HPDTLHKALQ-----DLINKIDQAHNEL----- 6650
QY 430 TGIWOHPAGAAR-----PQGESIR---LHDDKIHLHPGLGVQS 466

```

```

Db 6651 -----NGESRPFQALDNALNIDSINLSNLNVPQRTVKDNINHVTTLESQAEL----- 6698
QY 467 ADKDTSHQSLSRQADGKLYALKD-----NFTQLNLSDNKSEKLVDKIKSY--- 511
Db 6699 -----QKAKELNDAMKAWRDSIMNQEQIRKNSNVTNEDLAQONAVNHAVDKINNIIIGE 6751
QY 512 -----SVDQRGQVAILTDTPGPHKMSIMPSLD 538
Db 6752 DNATMDPQIITQATQDINTAINGLNGDQKIQDAKTDAKQOITNTFTGLTEPQKQAL----- 6806
QY 539 ASPESHISLSLHFAHAHQGLLHCK---SELEASQVSLSHGRLVAD-----SEGLKF 587
Db 6807 ---ENTINQOTSANVAKQLSHAKFLNGKMEELKVAVAKASLVRQNSYINEDVSEKEAY 6863
QY 588 SAAIPKQGGCNE-LKMKAMPQHALDEHFGHDHQSIGFFHDDHQQNLNALVKNFRQOHAACP 646
Db 6864 EQAIAK---GOEIIINSENNPTISSTDINRTIQEINDAEQNLHGD-----NKLRAQAEI- 6913
QY 647 LGNDHQFHPGWNLTDAVIDNQILGLHHTNPEPHEILDHGLSLALQEGKHLHFDQLTGK 706
Db 6914 AKNEIQNLGLNSAQITKLIQDIGRTTTPKPAVTKLEEAKAINQAMQOOLKQSIADK--- 6969
QY 707 WTGAESDCKQLKKGDLGAAVLLKDGVEVKRLNIN-----QSTSSIKHGTENVFSLPHVRN 760
Db 6970 -----DATLSSNVLNEDSEKKLAYDNVSAQAEQLINQNLNDPTWDISNIQAITQ 7018
QY 761 KP-EPGDALQGLNKKDKAQAAMAVIGVKNYLALTEKGD-----IRSFQIKPGTQQLERPA 813
Db 7019 KVIQAKDSLHGANKLAQNAQADSNLIINOSTNLNDKQKQALNDLNLHNAQTKQQAEEIIAQA 7078
QY 814 QTLISREGISGELKDIIHVD---HKQNLIALTHEGEVHFQPREAQWQNAESSW----- 862
Db 7079 NKLNNE--MGTLLKTLVEEQSNVHQSKY--INEDPQVQNIYNDSTQKREILNGTITDVLN 7135
QY 863 -HKLALPQSESCLKSLDMSHEHKPIATFEDGSQH-----QLKAGGWHAAYAAPERG 911
Db 7136 NKKIADAIQNIHLTKNDHGDQKLOKQAQDATNELNLTNLNNSQROSEHDEINSAERT 7195
QY 912 PLAVGTSGSQTIVNRLMGVKGKVIKPGSLTVKLSAQTGWTGAGRKVKSKFGERIRAY 971
Db 7196 EVSNDLNHAKAL-NEAMQLENEVALENSVK-KLSDFIN-----EDEAAQNEYSNALQKA 7248
QY 972 --AFNPWTST---PRPTKNAAYATQHWQCGREGKPLIYEMOGALIKOLD---AAN----- 1018
Db 7249 KDIINGVPSSTLDKATTEDALLEQNARESLHGQKLOEAKQNAVAEIDNLQALNPGQVL 7308
QY 1019 -----VRHNAPOPDLQSKLETLDLGEHGAELNDMKFRDELEQSASTRSVTVLGHQGV 1073
Db 7309 AEKTLVNOASTKPVQEQAL-----QKAKELNEAMKALTYEINKK-----EQIK 7351
QY 1074 KSGEINSEFPKPSFGKALVQSFN--VNRSGQ-----DLKSLQQAQVHAATPPSAESKL 1123
Db 7352 ADSRYVNAD-----SGLQANYSALNVGSQIIATTQPPPELNKDVINRATQITKTAENNL 7405
QY 1124 --QSMLGHFVSAG--VDMSHQGEIPLGRQDPNDKTALTAKSLRLDVTIIGELHELADKA 1180
Db 7406 NGOSKLAFAKSDGNQSTLEHQQ-----LTQSQK-----DKQHDNLNQA 7443
QY 1181 KLVSDHXPADQI---KQLRQOQFDTLREKRYESNPVKHYTD-----MGFTNKALEA 1229
Db 7444 QT---KQVDDIIVNNSKQLDNSMNQLQIIVNNDTVKQNSDFINEDSSQDAYNHAIOA 7499
QY 1230 NYDAVKAFINAFKKEH-----HGVNLITR----- 1253
Db 7500 AKDLITAHPTTMDKNQIDQAIENIKQALNDLHGSNKLSEKKEASEQQLNLSLNTNGQKD 7559
QY 1254 TVLSQSGSAEL-----AKKIKNTLLSL-----DSGESMGFSRSGYGGGVTFVPT 1298
Db 7560 TILNHIIPATRSQVEKIASAKQLNNTWKALRSIADNNEILQSSKYFNEDESEQ----- 7614
QY 1299 LSKKVPVPIPGAGITLDRAYNLSFRSTSGGLNVSFGRDGGVSGNMVATGHDVMPYMTG 1358
Db 7615 -----QNAYNQAVNKAKNIIIN-----DQPTPVMANDEIQSVLNE 7648

```

```

QY 1359 KKTSGN-----ASDWLSAKHISPDLRIGAAVSGTLQGTIONSLK-----1399
Db 7649 VKQTKDNLHGDKLANDKTDQAATNALNLYNAQGRNLETQVNSNSRPEVQKVQVLAN 7708
QY 1400 -----FKLTEDLPGEIHLGTHGLTPABLLQKGIHQMKQSGKLTFFSDTS--ANLD--1450
Db 7709 QNDAMKKLDDALTG-----NDAIKQTSNY-INEDTSQVQNFDEY 7747
QY 1451 LRAGINLNEDEGSKPNCVTVARVA-----GLSASANLAAGSRERSTTSQFGSTTS 1500
Db 7748 TDRGNIIVAEQTPNPNMSPNINTIADKITEAKNDLHGVOKLQAQOOSINTINQMTGLNQ 7807
QY 1501 ASN-----NRPTFLNGV--GAGANLTAALGYAHSTHEGKPVGIF 1538
Db 7808 AQKEQLNOEIQOTRSEVHQVINKAALNDMSMTLRQSIITDEHEVKQTSYINETVGNQ 7867
QY 1539 PAFSTNTVSAALANDNRSTQSIT--SLELKRAEPVTSN-DISELTSTLGHKHFQDSATTQML 1595
Db 7868 TAYNNA--VDRVKQIINQTSNPTMNPFLVERA--TSNVKISKDALHGERELNDNKNSTTF 7923
QY 1596 AA--LKELDADKAPBQLHLOQHF-----SAKQVVG-----1624
Db 7924 AVNHLNLAQAKKEALTHEIEQATIVSQVNNIYNKAKALNDMMKKLDLVAQODNVROSN 7983
QY 1625 -----DERYEAVERNKLKLVIROQAAD--SHSM 1649
Db 7984 NVINEDSTPQNMNDNTINHAQSIIDQVANPTMWSHDEIENAINIKHAI--NALDGEHL 8040
QY 1650 ELGSASHSTTYNNLSRIN--NDGIVELLHKKHFDALPASSAKRLGEMNNNDPALKDIIK 1706
Db 8041 QAKENANLLNSLNDLAPORDAINRLVNE-----AQTRKVAEQLQSAQALNDAMK 8093
QY 1707 QLQSTPFFSASVSMELK-----DGLREOTEKAILQGVREBVGVLFO-----DRNNLRV 1756
Db 8094 HLRNSIQNSSVRQESKYINASDAKKEQYNHAV-----REVENIINQHPFLDKKIIKQ 8147
QY 1757 KSVSVQSVKSGEGNTPALLLGTSNSAAMS-----MERNIGTINPKYQDQNTPRFRFLE 1812
Db 8148 LTDGVNQA--NNDLNGVELLDADKQNAHQSIPTLMHLNQAQONALNEKINNATRTVEVA 8204
QY 1813 GGIAQANQVASATLTKKEGLE 1835
Db 8205 AIIQOAK-----LLDHAVENLE 8221

RESULT 12
US-09-815-242-12610
; Sequence 12610, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

```

```

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12610
; LENGTH: 5795
; TYPE: PRF
; ORGANISM: Staphylococcus aureus
; US-09-815-242-12610

Query Match      2.7%; Score 257; DB 9; Length 5795;
Best Local Similarity 18.4%; Pred No. 8.3e-08;
Matches 377; Conservative 326; Mismatches 878; Indels 468; Gaps 91;

QY 16 TAAHNPVGHGVALQQ---GSSSSSPQNAASAALAEKGRKMPRIHQPSAADGISAHQ 72
Db 2647 TAAQRVINDGDATAQOISIDENTTAQRNDLTNQISQATNLAAVESVKQSANSLDG-AMGNL 2705
QY 73 QKKSFLRGCLGTTKXFSRPAQCGQGTTHSKGATLRLDLLARDGDTQHEAAAPDAARLTR 132
Db 2706 QTAINDKSGTTLASQNF--LDADQEKENAYNOAVSNAETILNKQTGPNTAKTAVEQALNNVN 2764
QY 133 S-----GGVKRRNMDMAGRPWKGS-----GEDKV-----PTQQRKHOL 168
Db 2765 SAKHALNGTQNLNAKQAIAITAINGASDLNKKQKDALKAQANGAQRVNSAQDVORNAETEL 2824
QY 169 NN-FQOMRQTMLSK---MAHPASANAGDRLOHS--PPHIFGSHHEIKKEP--VGSTSKATTA 222
Db 2825 NTAMGTLKHAIADKNTKLASSKYVNADSTKQNAVTKVTNAEHLISGTPVTVTSEVTA 2884
QY 223 HADRVETIAQEDDDSEFQOLHQORLARENPPOPKLGVATPI SARFQPKLTAVAESVLE 282
Db 2885 AANQVNSAKQELNGD-----ERLRAKQNAANTAI DALTQLTNPQAKLKEQVQGA--NRLE 2938
QY 283 GTDITQSPKLPQSM-----LKGSGAGVTPLAVTLDKGLQLAPDNPALNTLLKQTLG- 335
Db 2939 DIQTVQT--NQALNNAMKGLRDSIANET--TVKASQNYTASPNQSTYSAVSNAGKI 2994
QY 336 -----KDTQHYLAHASSDGSQHLLLDNKNKHLFDIKSTA-----TSYSLVHNSH 379
Db 2995 INQTNPTMDTSAITQATTQVNNAKNG-----LNGAENLRNAQNTAKQNLNTLSHLTNNQ 3049
QY 380 PGEIKGKLAQAGTGSVDGSKISLGSCTQSHKNTMLSQGEAHRSLLTGIIWOHPAGA 439
Db 3050 KSAISSQIDRAGHVSEVTAAKNAATEL--NTQMGN---LEQAIHQDQNTVKQGVNFTDADK 3104
QY 440 ARPQGESIRLHDDKIHILHPELGYVWQSAKDQTHSOLSQAQDGKLYALKDNRTLQNLSDNK 499
Db 3105 AKRDAYTNVSRSEAT--ILNKTQGA-NTSKQDVEAAIQNTVSATK-NALNGD---QNTVNAK 3158
QY 500 SSEK-LVDKIKSYSDVQGVAILTDTFGRHKMSIMPDLSDASPESHISLSLHFADAHQGL 558
Db 3159 NTAKHALLNLTSLINNAQKRDLTTKID-----QATTVAGVEAVSNTGTQLTAMANLQNGI 3213
QY 559 LHGKSELEAQSVALSHGRLVADSEK-LPSAAIP-----KQGDGNEKMKAMPQHALD 611
Db 3214 NDKANTLASENY---HD-----ADSDKKTAYTQAVTNAENILNKNSGSLD-KAAVAENALS 3265
QY 612 EHFCHDHQISGFPHDDHQLNA-----LVKNPFRQO-----642
Db 3266 QVTWAKGALNGNHLLEQAKSNANTTINGLOHLTTAQDKDLKQVQQAQNVAGVDTVTKSSA 3325
QY 643 -----HACPLGNHQQ-----FHPGMNLTAL-----VIDNOLGLHHTNPEPEILDM 684
Db 3326 NTLNGAMGTLRNSIQDNTATNNGQNYLDATESNKTNYNNAVDSANGVINATSNFN--MDA 3383
QY 685 GHLGSLALQSGKLYFDQLFKGWTGAESDCKQLK-----GLDGAAYL-----LKDG-----731
Db 3384 NAINQIATQVTS-----TKNALDGGTHNLTOAQKATNAIDGATNLNKAQKALKAQVT 3436

```



QY	684	MGLGLSLAQEGKLVHFDQLTKGWTGAESDCKQLKGLDGAAYLLKDGVEKRLNINOSTS	743
DB	1167	-----GTLASQ-----NFLD-----ADEQKRNAYNQVSAEETLN-----KQTGPNF	1207
QY	744	SIKHGTENVESLPHVRNKPFGDALQGLNKDDKAQAMAVIG---VVKYLALITEKGDIRSF	800
DB	1208	AVEQALNNVNNAKHALN-----GTQNLNNAKQAATAINGASDLNQKQKDALKAQANGA	1261
QY	801	QIKPGTCQOLRPACTLSREGISGELKDIHVDH-----KQMLY-----	837
DB	1262	QRVSNADQVOHNATELNT--AMGTLKHATAIDKTNNTLASSKYVNADSTKQNAYTTKVTNAE	1319
QY	838	-----ALTHEGEVEF-----HOPREAWQNGAES--SSWHKLALPQSES	872
DB	1320	HIISGTPVTVTTPSEVTAANQVNSAKQELNGDERLEAKQNAINTAIDALTQNTFQKAK	1379
QY	873	KLKSLDMSHEHKPIATFEDGSQHOLKA--GWHYAYAAPERGPLAVGTS-----GSQTV	923
DB	1380	LKEQVGQANLEEDVQVTQNGQALNNAMKGLRDSIANE---TTVKTSQNYTDAASPNNQST	143
QY	924	ENRLMOGVKGKVIPEGSLTVKLISAQTCGMTGABGRKVSFKSERIIPAYAFNPTMSTPRPI	983
DB	1437	YNSAVSNAKGIINQTNPTMOTSAITQATQVNNKAGLNGAENLR-----NAQNTAKONL	1492
QY	984	KNAAYATQHGWOEGREGLKPLYEMOGALIKQLD--AH---NVRHNAQPDPLOSKELTLDL	1037
DB	1493	NTLSHLTN-----QKSAISSQIDRAGHVSEVTATKNA--ATELNTQMGNEQ	1538
QY	1038	GEHGAEILLDMKRF-----RDELEQSATRSVTVLGQHGVLLKSGNEINSEFK--PSPG	1088
DB	1539	AJHDQNTVKQSVKFTDADKAKRDAYTNAVSRAEAILNKTQGANTSKQDVEAAIQNVSSAK	1598
QY	1089	KALVQSFNV-----NRSQDLSKSLQOAVHATPPSAESKQSLMGLCHF	1130
DB	1599	NALNGDQVNTNAKNAKNAALNMLTSINNAQKRLTTIKDQATTVAGVEAVNSTTQNTA	1658
QY	1131	VSAGVDMSHQKEIPLGRQRPNDKT-----ALTKSRLILD---	1166
DB	1659	MA-----NLQNGINDKTNLTASENYHDADSDKKTAYTAQVTAENILKNKS	1704
QY	1167	-----TVTIGELHELAD--KAKLVSDH-----KPDH-----DOIKQLRQO	1199
DB	1705	GSNLDKTAVENALSQVANAAGALNGHNLEQAKSNANTTINGLOHLTTAQDKQKQVQOQ	1764
QY	1200	-----FDTLREKRYESNPVKHYTDMGFTHNKALE-----ANY--DAVKA-----FINAF	1241
DB	1765	AQNVAGVDTVKSANTLNGA-----MGLIRNSIQNTATKQNGQNYLDATERNKTNYNNAV	1819
QY	1242	KKEHHGVNLTTRTVLSEQGSABELAKKLNLTLLSLDSGESMSFSRSYGGGVSTVFVPTLSK	1301
DB	1820	DSANGVINATSPNPDANAINQIATQVTSKXNALDGTNHL-----TQAK	1863
QY	1302	KYPVPVIPCAGITLDRAYN-----LFSRSTSGGLNVSPG-----	1335
DB	1864	QATNAIDGA--TNLNAKQAKALKAQVTSQARVANVTSIQOTANEINLTAMGQLQHGIDDEN	1922
QY	1336	-----RDGGVSGNI-----MVATGHDMVPMYTKTKTSAGNASDWL---SAKHKISPD	1379
DB	1923	ATKQTKYRDAEQSKKTAYDQAVAAAKAILNKQTSNDSKAAVDRALQOVTSKDALNGD	1982
QY	1380	LRIAGNAVSGTIQ--GTLQNSLKFKLTEDELPGFIHGLTHGTLTPAELLQKGIHEQMKGGS	1437
DB	1983	AKLAEAKAAAKQNLGTLNHTNAQRTD-----LEGQINQAT	2018
QY	1438	KL--TFSVDTSAN--LD-----LRAGIN-----LNEDGSKPQVTVARVSAGLSAS	1478
DB	2019	TVDGVNTVKTNTANTLDGAMNLSQGSINDKQATLRNQNYLDADESKRNYTAQVTA-----	2073
QY	1479	ANLAAGSRERSTTSGQFSGTTSASNRPFTLNGV--GAGANLTAALGVVAHSSTHEGKPVGI	1537
DB	2074	---AEGILNKQTG---GNTSKADVDA--LNAVTRAKAALNGADNLNRNAKTSATNTIDG	2124
QY	1538	FPAPTSTNVSAALDNRTSOSISLEKRAEPVTSISLELTSLTGLGHFKDSTATKMLA	1596

Db 2125 LPNLTQ-----LQKN-----LKHQVEQAQNVAGVGVKDKGNTL-----NTMAGALRT 2168  
QY 1597 ALKELDDAKPAEQHILQHFSAKQVGVDERVEAVNLKLVIRQQAADSHSMELGSASH 1656  
Db 2169 SIQNDNTTKTS-----QNYLDASDSNKNYNTAVNANGVI-----NATNPNNDANAING 2219  
QY 1657 STTYNLSRINDGIVELLHKHFDALPASSAKRIGEMMNDPALKDIIKQLQ-----1709  
Db 2220 MANQVNTTKAALNGAQLAQAKTNAINTINAHDLNOKQKD--ALKTVNNAQVSDANN 2277  
QY 1710 ----STFFSASVSMELDKGLRQT-----EKALDGKVGREBVGVLFODRNN 1753  
Db 2278 VQHTATELSAMTALKAAIAADKERTKASGNVYVADQEKRAQYDSKVTAENIIISGTPNAT 2337  
QY 1754 LRVKSVSVSQSVKSGFNTPALLLGTSNSAASMERNTGTFNFKYQDQONTFRRTLEG 1813  
Db 2338 LTVDVN-----SAASQVNAAKTALNGDNNLRVAKHEANNITD-----2375  
QY 1814 GIAQAN-----POVASALT 1827  
Db 2376 GLAQLNNAQKAKLEQVQSATT 2397

RESULT 14

US-10-282-122A-74463  
; Sequence 74463, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 74463  
; LENGTH: 2045  
; TYPE: PRT  
; ORGANISM: Streptococcus pyogenes  
US-10-282-122A-74463

Query Match 2.7%; Score 253; DB 12; Length 2045;  
Best Local Similarity 18.9%; Pred. No. 2.8e-08;  
Matches 418; Conservative 308; Mismatches 785; Indels 706; Gaps 100;  
QY 22 VHGVALQOQSSSS-----PQNAASALAAEGKNGKMPRIHOPSTAAADGISAHH 71  
Db 17 VGLGVAVPTGFGSQSGNMVVMVKAEEVATDLRSQASDSERVDSSLLQKENLSDV-----70  
QY 72 QQKKSFLRGCLGTTKKFSRPAQOGPGTTHSKGATLRDLLARDGGTQHEAAAPDAARLT 131  
Db 71 ----SFKLENLNGWEAENDTA--GNLGKFKDPDSSGVQNILTSSGKNISVAVAPK-----119  
QY 132 RSGGVKRRNMDDMAGRPVMKVGSGEDKVPQKQKHQNNP-----GQVRQTMLSKVAH 184  
Db 120 ----GSGKNIVKTKR--SNFGQGYVVGGLTQTPVLKL--152  
QY 185 PASANAGDRLQHSPPHPPGSHHEIKE--EVRGTSK-----ATTAAHADVET 229  
Db 153 ----NDVYRYSFTTKKLSGNSSEFKTRVFPVESNNKLGELVIRVNDKNVSTKHDMLPDI 208  
QY 230 AQEDDDSEFOQLHQORLARERENPPQPKLGATPISARFPQKLT-----274  
Db 209 SDGTHTVDFTLGDK-----KLSVA--FRSPRQTSNVVYFNSINIKNI 250  
QY 275 ----AVAESVLEGTDTTQSPKPKQSMKLGSGAGVTPLAVTLDKKG-----LOLAPDN 322  
Db 251 SPASVPAIPSKVLEGT-----SVLSGTAIS--SGDTLEKRSFGDILRVYKDS 297  
QY 323 PPAINTLLKQT-----LGKDTQHYLAHASSDQSHLLDNKGHLDIKS-----367  
Db 298 KIIARTVTKGNKNDVKLSKPLIAGEKLDPEILHPRSONVSKKISKQVEAFPPASVYKEK 357  
QY 368 ----TATSYSVLHNSHPGE-----IKGKLA--QAGTGSVSVQKSG 402  
Db 358 VIAKLFPVYEATSEKITNDAWLDENAKDLQOKLEQYISGKVAISEAGTKQEAIDAAYN 417  
QY 403 KI-----SLGSGTQSHNKTMLSQPG-----EA 424  
Db 418 KYSSQTDPSLPSQYKQNGKENEQEGRODLIOTRDLTLKAIQEDKWLTEQEKTIQEEA 477  
QY 425 HRSLLTGIV-----OHPAG-----AARPO 443  
Db 478 LKAFETGIESVNTVSLEQLKQRLIVYKASEKSEKKEYPEPINQHI PGKEKEVKAQK 537  
QY 444 GESIRLHDDKIHLHPELGVWQSAKDTSHLSQADGK--LYALKONRTLNL-----495  
Db 538 EELKCLHDTTLEKINQD--KWLTPDQAEQLQKQAEVTFKKQEAIKSAQTLTQLETDLAD 595  
QY 496 --SDNK--SSEKLVDKIKSYSDV---QRQVAILTDTGPRHKMSIMPSLDASPEHSLSL 549  
Db 596 YVSENEGKNSIPDKYKSGNKKDLVNAEYK--LKEAHEATKQATEKDPWLSPEOK-----649  
QY 550 HFADAHQGLLHGKSELEAQSVASHG--RLVVADS-----EGKLFSAAPKQ--594  
Db 650 ----KAQKEKAKARLDEGLKALKAAADSLEILKVTEAFVYDKKNPDSIPNQH 697  
QY 595 --GDGNELKMKAMPQHALDFHGHQISGFHDHGHQNLNVRKNFRQQHACPLGNHQ 652  
Db 698 KAGTADQARKQAL--DSLDK-----EVQKELESIDNDN-----TLTDEK 735  
QY 653 PFGWNLTDAIVDNLGLHHTNPEPHEILDGMHGLSALQEGKLYHFDQLTGKWTGAS 712  
Db 736 AAARKKVNDAVDVAKQTAMEANSYEDLTTIKDEFSLNLFHQKQ--TPLKDDQSDAIAELEK 794  
QY 713 DCKOLKGLDGAAYLLKDKGEVRLNINQSTSSIKHGTENVSFLPHVR-----NK 761  
Db 795 KQIEIKAEIGDKTILPRDEKEKQ--IADSKELASDKQKVDKADNADAKKAFEBEGKNI 852  
QY 762 PE---PGDALQGLNKDKQAQAMA-----VIGNVKYLALTE-----793  
Db 853 POAHIPGD-----LNK--DKEKLLAELKQKADDTKADTEKIDVDTKLTIDEKKEQKVTKAELEK 907





Db 1329 DVERLAQKQTAGNSLHLDLTTPAQQALENQINNATTRDKVA-----EIIA 1377  
 QY 276 VAESYLEGDTTQSPKFPQSM.KGSGAGVTPVLAUTLD--KGKQLAPD--NPPALNTLLK 331  
 Db 1378 QAQALNEAMKALKEGIDQDPQTEASSKFINEPQAKDQAYQVQAHAKDLINKTTPTLVK 1437  
 QY 332 QTLGKDTQHYLAHASSDSOHL-----LLDNKGHLFDIKSTATSYSLHNSHPCGE 382  
 Db 1438 SVIDQATQAVNDKKNLHGDQKLAQDKQATETLNLNSLNTFQQALENQINNAATRG 1497  
 QY 383 IKGKLAQAQAGSVSDGKSGKISLGSQTOSHNTKMLSPQCEAHRSLLTGIWHPAGAARP 442  
 Db 1498 VAQKLEA-----QALNQAM-----EALRNSIQOQOQTEAG----- 1528  
 QY 443 QESIRLHDKTHILHPELGVWQSA---DKDTHSQLSROADGKLYALKDNRTLNLNSDN- 498  
 Db 1529 ---SKFINEDK-----PQDAYQAAVQAHAKDLINQTSNPTLKAQVQELTQGVNQAKDL 1580  
 QY 499 KSSEKLVDKIKSYSDQRGQVAILTDPGRHKMSIMPSLDASPESHISLSLHPADA-HQG 557  
 Db 1581 HGDQKLADD-KQHAVTDLNQLNSL--NNPQQALESQIN--NAATREVAQKLABAQALDQA 1637  
 QY 558 LLHGKSELEAQAQVAISHGRILVADSEK-LFSAAIPOQGD-----GNELMKAMPQHAL- 610  
 Db 1638 MQALRNSIQOQOQTESSKFINEPKQDAYQAAVQAHAKDLINQTSNPTLKAQVQELTQ 1697  
 QY 611 -----DHFHGDHQ-----ISGFPHDGHQNLAL-----VKNNFROQHACP 646  
 Db 1698 AVTTAKDNLHGDQKLARDQOQAVTTYNALPNLHAQOQALTDAINAAPTKEVAQHVCQA 1757  
 QY 647 LGNDHQFHPGNLTDALVIDN-OLGLHHTNPEPEHILDMGHGLSALQEGKHYFQDLTK 705  
 Db 1758 TELDHAMETLKNVQVNTDKAQNTEASTDKKEAVDQ-----ALQAA-----ESITD 1806  
 QY 706 GWTGABSDCKQLKKGDLGAAAYLLKDGVEVKRLINQSTSSIK-HGTENVFSLPHVRNKPPEP 764  
 Db 1807 PTNGSNAN---KDAVEQALTKQE-KVNELNGNERVAEAKAQAKOTIDQLAHL-NADQI 1860  
 QY 765 GDALQGLNKDDKAQAMAVIGUNKYALATEKGDITRSFQIKGTQQLERPAOTLSREGISGE 824  
 Db 1861 ATAKQNTIDQATKLQPIAEL-VDAQTLNQSMQD---QLQAAVNEHANVEQTVDTYQADSD 1915  
 QY 825 LKDIHVDHKNLYALTHEGEVFHPREAWQNGAESSSWHKLALPQSBESKLKSLDMSHEHK 884  
 Db 1916 ---KONAY-----KQAAEAENVLKQNSNKKQ-- 1939  
 QY 885 PIATFEDGSHQKAGQWHAYAAPERGFLAVGTSGSOTV-----FNRLMQGVKKGK----- 934  
 Db 1940 ---VDQALQNLNAK---QALNGDERVALA-KINGKHDIDQLNALNNAQDGFGRIDQS 1992  
 QY 935 -----VTPGSLTVKLSAQTCGMGTGAER-----KYSSKFSERI----- 968  
 Db 1993 HDLNCIQOIVDEAKALNAMDQUSQIEISGNEGRTKGTSTNYVNAOTQVKQVDEADVAKAK 2052  
 QY 969 ---RAYAFNPTMTSPRTKNAAYATOHGWQREGKPLIYEMOGALIKQLDAHNVHRNAPQ 1025  
 Db 2053 ALDKSTGQNLTAQVILKINDAVTAAKKALNGEERLN---NRKSEALQRLDQLTHLNNNA-- 2107  
 QY 1026 PDLQSKLETDLGBHGAELINDMKRFRDELEQSATRSVTVLG-----OHQGVLSKNG 1077  
 Db 2108 ---QRQLAIQOI---NNAETLNKASRAINR---ATKLDNAMGAVQVQIDEQHLGVISSTN 2158  
 QY 1078 EINSB--FKPSPGKALVQSFNVNRSQGLSKLOQAQVHATPPSAESKLOSLMG--HFVSA 1133  
 Db 2159 YINADDNLKANYDNAIA-----NNAHELDKVVQNAI---AKABAEQLKQNIIDQALNALG 2210  
 QY 1134 GVDMSHOKGEI-----PLGRQRPNDKTKALTKSRLILDTVTIG-----ELHELADKAK 1181  
 Db 2211 DONLANAKDKANAFVNSLNGNLQOQDLAHNAINNADTVSDVTDIVNNQIDINDAMETLK 2270  
 QY 1182 -LVSDHKPDADQIKQLRQOQFDTLR-----EKRYESNPVKHYTDMGFTHNKALEANYDAV- 1234

2271 HLVNDNEIPNAQETVNYQNADDNAKTNFDDAKELANTLNSDNTVNDINGAIQAVKDAIQ 2330  
 1235 -----KAFINAFK-----KEHGWNLTRTRVLESQSA-ELAKKLXNTLLS 1274  
 2331 NLNGEORLOEAKDKAIQNVNKLKLEIEASNATDQDKLIJAKNAEELANSIINNINK 2390  
 1275 LDSGSMFSRSYGGGVSTVFPVTLSSKKVPVPIPGAGITLDRAYNLSFSRTSGGLNVSF 1334  
 2391 ATSNQDVSQVQTAGN-----QAIEQVHAHEIPKAKIDA---NKQVDKQVQALIDEI 2438  
 1335 GRDGGVSG-----NIMVATGH-DWPMYMTGKTSGNASDMLS----- 1371  
 2439 GRNPNTDKKEQALDKDRINQIILQOQHNDINNALT--KEAIEQAKERLAQALQIDKDLVKA 2496  
 1372 ---AKEHI-----SPDL-----RIGAAVSGTLOGTLONSLKFKLTED 1405  
 2497 KEDAKNKIKALANAKEDQINSNPDLTPEQAKAKALKEIDEAEKRALQ-NVENA---QTID 2551  
 1406 EL-PGPIHGL-----TH-----GTLTPAELLQKG--TEHOMKQSGKLTFSVD 1444  
 2552 QLNRLGNLGLDDIRNTHVWVEDDQPAVNEISBATPEQLLVNGELIVHR-----DD 2601  
 1445 TSANLDRAGINL-----NEDGSKPGVTVARSAGLSA--SANLAAGSR-----E 1487  
 2602 ITEQDVLIAHINLIDOLTAIEVIDTFS--TATISDSLTAKEVTEVTLDDGSKVIVNVPVKWE 2659  
 1488 RSTTSQFGSTTSASNNRPTFLNGVAGANLTAALGVAHSSTHEGKPVGIFPAFTSTNVS 1547  
 2660 KELSVMVQQAIESIENAVQOKINEINNSVTTLT-----EQK-----E 2696  
 1548 AALALDNRTSOSLSLELKRAEPVTS--NDISELTSTLGHKFKDSATTMKLA---ALKELDP 1603  
 2697 AATAEVNKLKQQAIDHINNAPDVHSVEEQOQEAHIEQFNPEQFTIEQAKSNAIKSIED 2756  
 1604 AKPAEQHLHLOQHSFSAKDVVGDREYAVRNKLKLVIRQQAAADSHSMELGSAHSTTYNNL 1663  
 2757 AIQ-----HMIDEIKARTDLTDKEQERAKLQNL--KEQA-----IQLIQAQSI----- 2800  
 1664 SRINNDGIVELLHKHFDALPASS-----AKRLGEMMNNDPALKDIIKQLQSTPFPSSAS 1717  
 2801 ---DEITEQL-EQFKAQMAANPTAKELAKRQEAISK---IKDFSNEKMAINSIRSEI 2851  
 1718 VSMELKDGLREOTEKAILDQKVGREEVGVLFODRNN-----LRVKSVSV- 1761  
 2852 GTADEKQAAMNQINEIVLE-----TIRDINNAHTLOQVEAALNNGIARISAVQIV 2901  
 1762 ---SQSVKSGEGFNTPAL--LGTNSAAMSMERNIGTINFKYQDQN 1804  
 2902 TSDRAKQSSSTGNSHSLTIGYGTANHPF-----NSSTIGHKKKLLDED 2945

Search completed: July 7, 2004, 15:30:44  
 Job time : 105.321 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:14:07 ; Search time 31.6095 Seconds

(without alignments)  
5593.254 Million cell updates/sec

Title: US-09-596-784-2  
Perfect score: 9448  
Sequence: 1 MELKSLGTEHKAHVTAHN.....NPQVASALTDLKEGLEMK 1838

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 T18448	pathogenicity factor
2	1824	19.3	1795	2 T30332	avirulence protein
3	275.5	2.9	6713	2 B89921	hypothetical prote
4	275	2.9	2514	2 F81045	hemagglutinin/hemo
5	258.5	2.7	2248	2 A35938	profilaggrin - hum
6	255.5	2.7	3455	2 B82519	hemagglutinin-like
7	247	2.6	3442	2 B82589	hemagglutinin-like
8	245	2.6	2703	2 H81193	hemagglutinin/hemo
9	244.5	2.6	3259	1 A56539	giantin - human
10	243.5	2.6	3225	1 E52300	giantin - human
11	241	2.6	2535	2 AC0304	probable hemolysin
12	240.5	2.5	5327	2 T13564	microtubule-associ
13	237.5	2.5	4152	2 T31102	filamentous hemagg
14	236	2.5	2271	2 F90073	hypothetical prote
15	232	2.5	1577	2 A35140	hemolysin A precu
16	229.5	2.4	2273	2 T09083	hemagglutinin/hemo
17	221.5	2.3	1635	2 A10452	hemolysin (importe
18	221	2.3	1608	2 A28182	hemolysin A - Serr
19	220.5	2.3	2253	2 T30336	nuclear/mitotic ap
20	220.5	2.3	3295	2 AE0074	probable adhesin Y
21	220	2.3	3890	2 C89921	hypothetical prote
22	219.5	2.3	2541	2 S11661	talin - mouse
23	218.5	2.3	4957	2 T03455	ALR protein - huma
24	218.5	2.3	5262	2 T03454	ALR protein - huma
25	217.5	2.3	1957	2 T38077	hypothetical coile
26	217	2.3	2712	2 T05113	hypothetical prote
27	214	2.3	3282	2 E82750	hemagglutinin-like
28	212.5	2.2	5627	2 C83339	hypothetical prote
29	210.5	2.2	1788	2 AH1447	probable tape-meas

30	210	2.2	2059	2 D82671	surface protein XF
31	209.5	2.2	1510	2 T31100	hypothetical prote
32	209.5	2.2	4919	2 T31105	hypothetical prote
33	207.5	2.2	3591	1 S21010	filamentous hemagg
34	205	2.2	1684	2 JW0057	gravin - human
35	204.5	2.2	4558	2 C82199	RTX toxin RtxA VCI
36	204	2.2	2845	2 I49505	adenomatous polyo
37	203.5	2.2	1995	2 G81044	hemagglutinin/hemo
38	203.5	2.2	3187	2 JC5837	364K Golgi complex
39	202.5	2.1	2218	2 B84683	hypothetical prote
40	202	2.1	1975	2 B81192	hemagglutinin/hemo
41	200	2.1	1467	2 A75564	conserved hypochet
42	199	2.1	2232	2 T34434	hypothetical prote
43	198.5	2.1	835	1 I57441	involutacin - orang
44	197.5	2.1	2015	2 B81989	hypothetical prote
45	196.5	2.1	1902	1 B44858	lactocepin (EC 3.4

## ALIGNMENTS

### RESULT 1

T18448  
Pathogenicity factor DspA - Erwinia amylovora  
C:Species: Erwinia amylovora  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T18448  
R:Gaudriault, S.  
submitted to the EMBL Data Library, May 1998  
A:Reference number: Z18936  
A:Accession: T18448  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1838 <GAU>  
A:Cross-references: EMBL:Y13831; PIDN:CAA74156.1  
C:Genetics:  
A:Note: dspa  
C:Function:  
A:Description: involved in pathogenicity

Query Match 100.0%; Score 9448; DB 2; Length 1838;					
Best Local Similarity 100.0%; Pred. No. 0;					
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1	MELKSLGTEHKAHVTAHN	PVGHGVALQGGSSSSPQNAASLAAGKNCMPRIHQP	60	
Db	1	MELKSLGTEHKAHVTAHN	PVGHGVALQGGSSSSPQNAASLAAGKNCMPRIHQP	60	
Qy	61	STAADGISAAHOOKSFSLRGCLGKTKFSRSAPQGPQGTTHSKGATRLDLLARDGGETQH	120		
Db	61	STAADGISAAHOOKSFSLRGCLGKTKFSRSAPQGPQGTTHSKGATRLDLLARDGGETQH	120		
Qy	121	EAAAPDAARLTRSGGVKRENMDMAGRPMVKGSGEDKVPTQOKRHLNFGQMTLS	180		
Db	121	EAAAPDAARLTRSGGVKRENMDMAGRPMVKGSGEDKVPTQOKRHLNFGQMTLS	180		
Qy	181	KMAHPASANAGRLQHSPPHIFGSHHEIKKEEPVGSSTKATTAHADRAVEIAQEDDSEFQQ	240		
Db	181	KMAHPASANAGRLQHSPPHIFGSHHEIKKEEPVGSSTKATTAHADRAVEIAQEDDSEFQQ	240		
Qy	241	LHQORLARENPPOPPLKGVATPISARFQPKLTAVAESVLEGTDTTQGPLKQSMKGS	300		
Db	241	LHQORLARENPPOPPLKGVATPISARFQPKLTAVAESVLEGTDTTQGPLKQSMKGS	300		
Qy	301	GAGVTVLAVTLDKGKLQALPDNPALNTLLKQTLGKDTQHYLAHASSGSGSHLLDNKG	360		
Db	301	GAGVTVLAVTLDKGKLQALPDNPALNTLLKQTLGKDTQHYLAHASSGSGSHLLDNKG	360		
Qy	361	HLFDIKSTATSYSLVHNHSHPEIKGLAQAGTGSVSDGKSGKISLGSCTQSNKTMLSQ	420		
Db	361	HLFDIKSTATSYSLVHNHSHPEIKGLAQAGTGSVSDGKSGKISLGSCTQSNKTMLSQ	420		
Qy	421	PGENHRSLLTGIWQHPAGARPQGESIRLHDDKTHILHPGLVWQSAKDQTHSLSRQAD	480		

Db 421 PGEAHSLLTGTWHPAGARFQGESIRLHDDKHILHPELGWQADKDTTHSQLSQAD 480  
QY 481 GKLYALKONRTTQNTSDNKSSEKLVKISYSVDQGOVAILTDTTPGRHKMSIMPSLSDAS 540  
Db 481 GKLYALKONRTTQNTSDNKSSEKLVKISYSVDQGOVAILTDTTPGRHKMSIMPSLSDAS 540  
QY 541 PSHSISLSHFADAHQCLLHGKSELBAQSVASHGRLVVADSEGLFSAAIKPKQGDGDEL 600  
Db 541 PSHSISLSHFADAHQCLLHGKSELBAQSVASHGRLVVADSEGLFSAAIKPKQGDGDEL 600  
QY 601 KMKAMPQHALDRHFHGHDIQSGFFHDDHGLNALVKNPNRQOQHACPLGNDHGFHPCWNLT 660  
Db 601 KMKAMPQHALDRHFHGHDIQSGFFHDDHGLNALVKNPNRQOQHACPLGNDHGFHPCWNLT 660  
QY 661 DALVIDNOLGHLHTNPPEHEILLDMHGLSLALQEGKLYHFDQLTGWTGAESDCKQLKKG 720  
Db 661 DALVIDNOLGHLHTNPPEHEILLDMHGLSLALQEGKLYHFDQLTGWTGAESDCKQLKKG 720  
QY 721 LDGAAYLLKDGVEKELNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDKAQAM 780  
Db 721 LDGAAYLLKDGVEKELNINQSTSSIKHGTENVFSLPHVRNKPPEGDALQGLNKDKAQAM 780  
QY 781 AVIGWNYKLYALTEKGDIRSFOIKPQTQOLERPQATLSREGISGELKDIHVDHKQNLAYLT 840  
Db 781 AVIGWNYKLYALTEKGDIRSFOIKPQTQOLERPQATLSREGISGELKDIHVDHKQNLAYLT 840  
QY 841 HEGEVFHPRAWONGAESSWHKALPQSESKLSLSDMSHEHKPIATFEDGSHQHLKAG 900  
Db 841 HEGEVFHPRAWONGAESSWHKALPQSESKLSLSDMSHEHKPIATFEDGSHQHLKAG 900  
QY 901 GWHAYAAPERGLAVGTSGSTVFNRLMOGVKGKVPISGSLTVKLSAQTGMTGAERKV 960  
Db 901 GWHAYAAPERGLAVGTSGSTVFNRLMOGVKGKVPISGSLTVKLSAQTGMTGAERKV 960  
QY 961 SSKFSERIRAYAFNPTMSTRPINKAAYATQHGQREGKPLKPLYEMOGALIKQIDAHNVR 1020  
Db 961 SSKFSERIRAYAFNPTMSTRPINKAAYATQHGQREGKPLKPLYEMOGALIKQIDAHNVR 1020  
QY 1021 HNAPOPDLOKLETLIDLGEGAEILLNDMKRFRDELEQSASTRSVTVLGHQGVLSKNGEIN 1080  
Db 1021 HNAPOPDLOKLETLIDLGEGAEILLNDMKRFRDELEQSASTRSVTVLGHQGVLSKNGEIN 1080  
QY 1081 SEKPSFGKALVQSENNRSGQDLKSLOAVHATPPSAESKQSLMGLHFVSVAGVMSHQ 1140  
Db 1081 SEKPSFGKALVQSENNRSGQDLKSLOAVHATPPSAESKQSLMGLHFVSVAGVMSHQ 1140  
QY 1141 KGBIPLGRQDPNDKTALTKSRLILDTVTIGELHELADKAKLYSDHKPDADQIKQROOF 1200  
Db 1141 KGBIPLGRQDPNDKTALTKSRLILDTVTIGELHELADKAKLYSDHKPDADQIKQROOF 1200  
QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTTRVLESQ 1260  
Db 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKKEHGVNLTTTRVLESQ 1260  
QY 1261 SAEALAKKNTLLSLDGSMSFSRSYGGVSVTFVPTLSKKVPVPVPCAGITLDRAIN 1320  
Db 1261 SAEALAKKNTLLSLDGSMSFSRSYGGVSVTFVPTLSKKVPVPVPCAGITLDRAIN 1320  
QY 1321 LFSFRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380  
Db 1321 LFSFRTSGGLNVSFGRDGGVSGNIMVATGHDVMPYMTGKTSAGNASDMLSAKHKISPD 1380  
QY 1381 RIGAAVSGTLOQTIONSLKFKLTDELPFGFIHGLTHGTLTPAELLQKGIHOMQGSKIT 1440  
Db 1381 RIGAAVSGTLOQTIONSLKFKLTDELPFGFIHGLTHGTLTPAELLQKGIHOMQGSKIT 1440  
QY 1441 FSVDTISANLDRAGINLNDGSKPNGTARVSGAGISASANLAAGSRERTTSGQGSITS 1500  
Db 1441 FSVDTISANLDRAGINLNDGSKPNGTARVSGAGISASANLAAGSRERTTSGQGSITS 1500  
QY 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560

Db 1501 ASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVGIFPAFTSTNVSAALALDNRTSQSI 1560  
QY 1561 SLELKRAEPTVNDISELTSTLGKPKDSATTMMLAALKELDDAKPAEQHILHQHFSK 1620  
Db 1561 SLELKRAEPTVNDISELTSTLGKPKDSATTMMLAALKELDDAKPAEQHILHQHFSK 1620  
QY 1621 DVVGDERVEAVNLKGLVIRQQAADSHSMELSGASHSTTYNNLSRINDGIVELLHKHFD 1680  
Db 1621 DVVGDERVEAVNLKGLVIRQQAADSHSMELSGASHSTTYNNLSRINDGIVELLHKHFD 1680  
QY 1681 AALPASSAKRLGEMNNNDPALXDIKQLOSTPEFSSASVSMELKDGRLREOTEKAILDGKVG 1740  
Db 1681 AALPASSAKRLGEMNNNDPALXDIKQLOSTPEFSSASVSMELKDGRLREOTEKAILDGKVG 1740  
QY 1741 REEVGVLFQDRNRLRVKSVSQSVKSEGFNTPALLLGTSNSAAMSERNITGTFNKYG 1800  
Db 1741 REEVGVLFQDRNRLRVKSVSQSVKSEGFNTPALLLGTSNSAAMSERNITGTFNKYG 1800  
QY 1801 QDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838  
Db 1801 QDQNTPRFTLEGGIAQANPOVASALTDLKKEGLEMK 1838

## RESULT 2

T30332

C:Species: Pseudomonas syringae

C:Date: 22-Oct-1999 #revision 22-Oct-1999 #text\_change 22-Oct-1999

C:Accession: T30332

R:Bogdanove, A.J.; Kim, J.F.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.;

Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998

A:Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dspeF

A:Reference number: Z20825; MUID:98115919; PMID:9448330

A:Accession: T30332

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1795 &lt;BOG&gt;

A:Cross-references: EMBL:U97505; NID:g2978502; PID:g2978503; PIDN:AAC06134.1

C:Genetics:

A:Gene: avrE

Query Match 19.3%; Score 1824; DB 2; Length 1795;

Best Local Similarity 29.0%; Pred. No. 8.9e-83;

Matches 555; Conservative 366; Mismatches 744; Indels 246; Gaps 72;

QY 55 PRH-----QSTADGISAHQKKSFSLGCLGKFKFSRSPQGFTHSKGA-- 105

Db 4 PSIHRTGSIQPTVTPDARAATDLOER-----AEQPRQRSSHSLSVG 47

QY 106 --TLRDL--LARDGGETOHEAAPDAAR-----LTRSGGVKRRNMD-- 143

Db 48 KRALKSVGLFKQSKAPQQAATPTAKNVKTPPPASNVATPRNKARESGFSNSPONT 107

QY 144 -----MAGRPYKGGSGEDKVPTQQRHOLNFGQWOTMLSKMAHPASANAGDRLQHS 197

Db 108 RAPKWLIRNHPQAQSSG-----AQTHEIPEAAPRKNLVRFDLPQ-----DLERS 155

QY 198 PHIPGSHHEIKBBPVGTSKATTAHADREVEIAQEDDDSEFQOLHQRLARENPQP- 256

Db 156 PSYSDNDPMTDEEAVANATROF-----RSPDHLQSGDGTIRMLATDPDQS 204

QY 257 ---PKLGAT-PISARFQPLTAVAESVLEGTDTTQSPKQPSMLKGSAGVTP-LAVTL 311

Db 205 SSGSKIGSDGPIPR-----EPMWRSNGGRFELKDEKIVRNS-----EPOGSIOL 251

QY 312 D-KGKLQAPNPPALNTLLKQTLGKDTQHYLAHASSDGSQHLLDNKGHLFDTKSTAT 370

Db 252 DAKGKPDFSTFPGAPLLDSILATPKQTYLAHQSKGVHGHQQLQANGHLHLAQDSS 311

QY 371 SVSVLHNSHPG-EIKGKLAQAGTSVVDGSKGKISLGSCTGQSHNKTMLSQPGEAHRSL 429

Db 312 SLAVIRSNALLIEGKKPPA-----VKMEREDGNTHI--DTASGRKTQ-ELPGKAHIAH 364

QY 430 TGIWHPAGARPGESIRLDDKHILHPGLGWOSAD--KDT-HSLSRQADGKLYAL 486  
 Db 365 TNV-----LLSHDGERMRVHEORLYQFDPISTRWKIPEGELEDTAFNSLSTGGNGSVYAK 418  
 QY 487 KDNRTLONLSDNKSSEKLVKIKSYSDQGOVAILLTDPTRHKMSIMPSLDASPEHSIS 546  
 Db 419 SDDAVV-DLSSPMPHPHYVEDLQSFVAPDNRAALLS---GKTTQAILLT-DMSPP---V 469  
 QY 547 LSHFADAHQGLLHGKSELBAQSAVAISHGRILVADSEGLFSAALPKQGGNGLKMKAMP 606  
 Db 470 IGGTLPTKKTGLELDGGKAQAAVGLSGDKLFIADTQRLYSAD-RSAPEGDDPKLKMPP 528  
 QY 607 QHALDE---HFHGDHOLSEFHDHQLNALVKNNFRQHQACPLG-NDHQFHPGWNLTLD 661  
 Db 529 EQANFQLEGVPLGHNRTVGFINGDDGVGVALIKNRQGETHSHALDEQSSKLQSGWNLTN 588  
 QY 662 ALVIDNOLGLHTTNEP---HEILDMHGLSLALQEGKLYHFDQITKGWTAEB-SDKQL 717  
 Db 589 ALVLNNRGL--TWPPPTAADRLNDRAGLVGLSEGRQIQRWDATPECWKDAGIKDIDL 646  
 QY 718 KGLDGAAYLLKDEVKRLNINQSTSSIKHGTENVFSLPHVYRNKPEPGDALQGLKODKA 777  
 Db 647 QRGADSNAYVLKGLKHALKTAABHPNNAFDRNTALATARTKVMGKEIEGL--DDR 704  
 QY 778 -QAMAVIGVNVKYLALTEKGDIRSFOIKPQTQQLERPAQTLREGISGBELKDIHVDHKONL 836  
 Db 705 IKAFAMVSNKRFVALDD-----QNKLTASHKOHKPVTLIDPGLGDKSLSLDEKHL 757  
 QY 837 YALTHEGEVFPQPRAWQN---GAB-SSSWKHALPQSEKLSLDMHSEHKPIATFDDG 892  
 Db 758 HALTSTGGLYCLPKAWQSTKGLDQLRARWTPVALPGQG-PVKALFTNDNDVLSAQIEDA 816  
 QY 893 SOH-----QLKAGWHAYAPERGFLAVGTSGSQTVENRLMQGVKGVIPGSGTLVKLSAQ 949  
 Db 817 ECKGLMQLKAGWQRF---EORP--VEENGLNDVHSRTGSKNTWRIPKIGITLMDVNT 871  
 QY 950 GQMTAGRGKVSSEKSERIRAVAFNPTMTSPRIKNAAYATQHGQWQREGKPLXLYEMQQA 1009  
 Db 872 FORSGVEKSKAST-SEFIRANIYKNTAETPRWKNVGDHQRVQGRGLKEVYETESM 930  
 QY 1010 LKQLDAHNVRHNPQP-----DLSQKLETL- -LGEHGAELNLMKPRDRLEQSATRS 1062  
 Db 931 LFKQLEL--IHESGGRPPARGQDLKARITALEAKLGPQGATLVKELETLDLENHSYTA 988  
 QY 1063 VTVLG-----OHQGVLSNGEINSEFPSP-----GKALVQ--SFVNNRSGODLS 1105  
 Db 989 LMSIGSGYKAKNLKQODGILNQHGL-----AKPSVMQFQKGLADLTGTLNFKSSGHDV 1045  
 QY 1106 KSLQQAHVHATPPSAESKLSMLGHFVSAGVDMSHQGEIPLGRQRPNDKTLTKSLRL 1165  
 Db 1046 KELQDALTOVAPSAENPTKLLGLTLKHQGLKSHQKADIPLGQRDASEDHGLSKARLAL 1105  
 QY 1166 DVTIIGELHELADKAKLVSDHKPDADQIKQLRQOFTLREKRYESNPVHYITDMGTHNK 1225  
 Db 1106 DLVTLKSLGALLDQV---EQLPPQSDIEPLQKLAITLDRVTVGENPVKVTWDMGFTDNK 1161  
 QY 1226 ALEANYDAKAFINAKKEHGVNLTRTVLESQGSALAKLKNLTLLSLDSG-BSMSPS 1284  
 Db 1162 ALESYESVKTFKLSFKKADHAVSVNNRAATGSKQDAELAGFKMLKQLEHGDDEVLQ 1221  
 QY 1285 RSYGGGVSTVFVPTLSKVPVPIPGAGITLDRAYNLSFSRSTSGGLN---VSFRGRDGGVS 1341  
 Db 1222 RSYGVNLTTFIILADK--ATGLWPTAGATGNRYIILNAERCEGGVTLYLISEGA-GNVS 1278  
 QY 1342 GNMIVATGHDVMPYM-----TGKTSAGNASDMLSAKHKISPDIRIGAAVSGTLQTLQN 1396  
 Db 1279 GGF--GAGKDYWPGFDPANNPARSDVGN-----NRTLTNFRGLVDVTVTAASQRA 1329  
 QY 1397 SLKFKLTELDPFIHGLTHGLTTPAELLQKIEHQMOKSKLTFSDVTSANLDRAGIN 1456  
 Db 1330 GVVFNVPDEIDAFVDDLFEQLMPLQVLKAVDYESYARRNFNFDLTAGTADIRAGIN 1389  
 QY 1457 LNED--GSKPN--GVTVARVSAGLSA--SANAAGSRERSTTSQGFSTTSASNNRPTFL 1509

Db 1390 LTEDRDPNADPNSDSFSAVVRGGAFAANITVNLMTYTDYSLTKNDKTELKEGKNRREFL 1449  
 QY 1510 NGVGAGANLTAALGVAAHSSTHEKPV---GIFPA--FTSTNVSAAL--ALDNRTSQSISL 1562  
 Db 1450 NNTVAGQLRAQIGGSHAP-TGTPASAPGPTPASQTAANNLGGALNFSVENRTVKRIKF 1508  
 QY 1563 ELKRAEPTVNSDISELTSTLGKHFQKFSATTYMLAALKELDDA-----KPAE---QLHIL 1613  
 Db 1509 RYNVAKBITTEGLSKLSGLGEAFLODNTTKAKLAELADPLNARYTGKKPDEVIAQDGL 1568  
 QY 1614 QOHFS-----AKVVGDERYEAVRNKLKLVIRQQA-ADSHSMELGSGASHSTTYNNLSRNN 1668  
 Db 1569 EELFADIPPPKD--NDKQYKALRDLKRAAVEHRASANKHSV-MDNARFETSKNLSGLSS 1625  
 QY 1669 DGIV-ELLKHFFDAALPASSAKRLGEMNNDPALKDIIKQLOSTPFFSASVSMELKGLR 1727  
 Db 1626 ESILTKIMSSVRDASAP-GNATRVAEFMQDPKRLAMLKEMEGSIGTILARVLEPKDSL 1684  
 QY 1728 EOTEKAILDGKVGREEVGVLPQDRNNLKVSVSVSQSVSKSEGFENTPALLLGTSNSAAMS 1787  
 Db 1685 DKIDEGSLNGTMTQSDLSMLEDNEMRIKRLVVFHTATQAEFTSPPTPLSYNSGANVS 1744  
 QY 1788 MERNIGTINFKYGODQNTPRFTLEGGIAQANPOVASALITDLKKEGLEMK 1838  
 Db 1745 VTKTLGRINFGVAGDQDKPIGTTFDGELSRPSASLKEAAGDLKKEGFELKS 1795

RESULT 3  
 B89921  
 Hypochemical protein ebha [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: B89921  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc, ma, A.; Mizutani-Di, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; PMID:21311952; PMID:11418146  
 A:Accession: B89921  
 A>Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-6713 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BA842527.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: ebha

Query Match 2.9%; Score 275.5; DB 2; Length 6713;  
 Best Local Similarity 17.8%; Pred. No. 0.00021;  
 Matches 407; Conservative 322; Mismatches 904; Indels 651; Gaps 95;

QY 11 KAAVHTAAHPVGHGVAL-----QOGSSSSSPQAAAAAAGNKGKMPRIHQFSTA 63  
 Db 556 KAAVENALSQVTNAKALNHNLEBAKSNANTINGLOHLITTAOKDKLKQ-QVQAQNV 614  
 QY 64 ADGISAHHQKKSFSIRGCLGKTKFSRPAQOQPGTHHSGKATRLDLLADGETOHEAA 123  
 Db 615 A-GVDIV--KSSANTLINGAMGTLRNSI-----QDNTATKNGQNYLD--ATERKNTYNNNA 664  
 QY 124 APDAARLTSGGVKRRNMDMAGRPMVKGSGEDKVPQO--QKRHOL---NNFGOMRQTM 178  
 Db 665 VDSANGVINA--TSNPNMD-----ANAINQATQVTSKGNALDGNLTQAKQTA 712  
 QY 179 LSKMAHPASANAGDRLQHPHPPHPSGSHHEIKPEVPGSTSKATTAAHADV-----EIAQ 231  
 Db 713 TNAIGATNLNKAQK-----DALKAQVTSAGRANVTISIQTAN 751  
 QY 232 E-----DDSEFQQLHQORLARENPFPKLGVAPTSAREPQKLT--AVA 277  
 Db 752 ELNTAMQQLQHGIDDENATKQTKYRDAEQSKKTAQVAAAKAILNKQTSNSDKAAV 811



Science 287, 1809-1815, 2000  
 A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
 A;Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.  
 A;Reference number: A81000; MUID:20175755; PMID:10710307  
 A;Accession: F81045  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-2514 <TEXT>  
 A;Cross-references: GB:AE002026; GB:AE002098; NID:g7227015; PIDN:AAF42109.1; PID:g722702  
 A;Experimental source: serogroup B, strain MC58  
 C;Genetics:  
 A;Gene: NMB1768

Query Match 2.9%; Score 275; DB 2; Length 2514;  
 Best Local Similarity 18.8%; Pred. No. 4.3e-05;  
 Matches 389; Conservative 287; Mismatches 803; Indels 586; Gaps 91;

QY	7	GTEH-KAAVHTAAHPV-----GHCVALQOQSSS-----SSPQNAAS 43
Db	37	GSADVKSVPFGTTHAPVCRSNIFSFSLGFLSLAVGTANTAFADGIIADKAAPTQAT 96
QY	44	LAAPGKRGKMRHQPSTADGSAHQAQKKSFLRGCLGKTKKFSRQPG-----98
Db	97	ILQTGNG---IPQVNIQTPTSAGSVNVQYACFDVGNRGAILNN--SRNTQTQLGGMIOG 151
QY	99	-----TTHSK-----GATLRDLARDDGETOHEAAAPDAARLTR 132
Db	152	NPWLARGEARVVVQINSSSHSSQNGYIEVGRRAEVVIANPAGIANGGFINASRATL 211
QY	133	SGGVKRRNMDMAGPMVKG-----GSGED-----KVPTQQRHQLNFGQMRQTMLS 180
Db	212	TGQPYQAGLSEGFKPIQGNVVIAGHGLDARDTDFILSVHSKIDAPVVGQDVRVAVG 271
QY	181	KMAHPASANAGDRLOHSPHPIPGSHETKEBEPVSTKATTAHADRVIAQEDDDSEFQQ 240
Db	272	QNDVVATGNA---HSP-----ILNNAANTSNNTANNTHPIPLFADTG-----312
QY	241	LHQRLARERENPPQPKLGIVATPISARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300
Db	313	-----KLQ-----GMVAKITILISTA-----BOGIRNQQLPAS 342
QY	301	GAGVTPLATVLDKGLQAPNDPAPALNTLLKQTLGKDTQHYLAHAS--SDGSHLLLDN 358
Db	343	SGN-----VAIDANGLVNSGTMAAN-----AKDONTAEHKVNIERSQG-----VEN 385
QY	359	KG-----HLFDIKSTATSYS-----VLNHPGEIKGLAQAGTGSVSDGSKGI 404
Db	386	SGTAVSQQGTQIHSQSIQNTGTLSSGELIHN-----GSLKNETSCTI-----EAARL 435
QY	405	SLGSGTQSHNKTWLSPQGEAHRSLLTGIWHPAGARPQGESIRLHDDKIHILPELGVW 464
Db	436	AIDTDT-LNNQKLSQTG-----SQKLI-----458
QY	465	QSADKTHSOLSRQADKGLYALKDNRTLQNLSD-----NKSSEKLVDKIKSYVDORQQ 518
Db	459	-----DAQKM-----DNGRMGLQDTAPTASNGSNQGN-----SYNASEHSS 498
QY	519	VAILTDPGRHKMSIMPSLDAPESHIISLHPADAHQGLHGKSELEAQSVASHGRILV 578
Db	499	TTTTPTATGTGTATVSISNITAPT-----FAD---GTIRTHGALDNGSGSIANGQTD 547
QY	579	VADSEKLSAALPKQGDGNELKMKAMPQHALDEHFGHDHQISGFFDHDGQLNALVKN- 637
Db	548	VSAQQG-LNNAQ---QIDIHQLNAKG---SAFDNNG--TIIISDAVHIQAGSLNNGNI 598
QY	638	NFQQQACPLGNDHQHPGWNLTALVID-NQLGLHHTNPEPEHILDMHGLSLALQEGK 696
Db	599	TRQQ-----LEIETDQLDNAHGKLLSAEIALDVLAVSGSLNNGNQE 638
QY	697	LHYFDQLTKGWTGAESDCKOLKKGLDGAAYLLKDGVEKVRININQSTSSIKHGTENFSLP 756
Db	639	IATNQQL-----IHDGQOSTAVIDNTNGTIQSGRDVAIQAK 675

QY	757	HVANKPEPGDALQGLNK-----DDKAQAMAVIGVKNKYLTALTEKGDIRSFOIKPGTQOLE 810
Db	676	SLSN-----NGTLAADNKLIDIALQDDFYVERNIYAGNE-LSLSTRGSLKUNSHLTQAGRRIR 730
QY	811	RPAQTLISREG-----ISGELDIHVHDKONLYALTHEGEVHFHQPREAWQNGAESSESMHKL 866
Db	731	IKANNLDNAAQNIQSGGTTDGTQHN-----LTNRGLIDGQ-----767
QY	867	LPOSESKLSLDMSH-----EHPFIATFEDSGSHQLKAGGWHAVAAEPGRPLAYGTS 918
Db	768	-----QTKIQAGQMNIGTGTIGRIYGNIAIAATALDNDQDENGTA--ATAARENLMGLIG-- 819
QY	919	GSQTVFNRLMOGVKVIPOGSLTVKLSAQTGGMTGAEGRKVSKFSERIRAYAFNPIMS 978
Db	820	-----QLNRENLSIYSGNDMAVGALDNTGQATKAQR-----853
QY	979	TPRIKAAVATQHGMOGREGKLPYEMOGALIKQL-----DAHNVNRNAP 1024
Db	854	-----IHNAGATIEAAGKWRILGVEKLNHTNHLKTLQVETGREHIVDYEAFGREHLLREGT 909
QY	1025	QPOL-----QSKLETLDLGEH-----GAEILLDMK 1049
Db	910	QHELGSVYNDESDDLRTPDGAAHENMHKYDEYKVTQKTQVTPAPAKIISGNDLTDGK 969
QY	1050	R-FRDELEQSATRSVTVLGGHQ-----VLKSGEINSEFKPS-PGKALVQSFN 1096
Db	970	EVFNTDSQIITAGGNLIYQTEKDXGLHNEQTGCEKKVFBENGKLSHYWREKHKGR---DSTG 1026
QY	1097	VNRSGDLSKLSQAVHATPPSAESKLQSMGLHFVSAGVDMSHQGEIPLGRQRPDNDKT 1156
Db	1027	HSEQNYTLPEBITRNI SLGSPAYESHKALSHHAPSQGTETLPQSG---ISLPTSNSFT 1083
QY	1157	ALTKSRL-ILDTVTITIGELHELADKAK-----LVSDHKPDA--DOIROLROQDFTLREKR 1207
Db	1084	PLPSSLYIINPVNKGVLVETDPRFANYRQWLGSDYMLDSLKLPNNLHKLGLDGYEQR 1143
QY	1208	YESNPVHYTDMGTNKALEANYDAKAFINAKKEHHGVNLITRTVLESGQSALAKK 1267
Db	1144	LINEQIAELT--GHRRLDGYQNDDEQFKALMDNGATAAARSNNLSVGIAL-----SAGVQAQ 1197
QY	1268	LKNTLLSLDSGESFSRSYGGVSTVFVPTLSKKVPVPVPIPGAGITLDRAYNLSFRSTS 1327
Db	1198	LTSDIVLWLVQKE-----VKLPDGGTQTVLPQVYVVKNGDIDGKAL-----LGSNTQ 1247
QY	1328	GGLNVSFGRDGVSGNIMVATGHDVMPYMTGK---KTSA-----GNASDWLSAKHKIS 1377
Db	1248	INVSGLKNSGTIAGRNALLINTDLDNIGGRIHAKSAVATATQDINNIGGMLSAEQTL 1307
QY	1378	PDLRTGAAY-SGTLOGTQNSLKFKLTEDELPG-FIHGLTHGTLT-----PAELLQKGIE 1430
Db	1308	--LNAGNNINSQSTTASSONTQSSSTYLDRMAGIYITGKEGVLAQAQKDNIIAQGIS 1365
QY	1431	HQMKQGSKLTPSVDTSANLIDL-----RAGINLNEGSKPENGVTARVSAGLSASA-----1479
Db	1366	NOSEQG-QTRLQAGRDINDTQVTSKHQTHFDADNHVIRGSTNEVSGSIQIKGDTVLLS 1424
QY	1480	--NLAAGSERSTTSGFGSTTSASNRPFLNGVGAGANLTAALGVAHSSTHEGKPVGI 1537
Db	1425	GNNLAKAAEVSSANGTL--AVSAKND---IN-ISAGINTT---HVDDASKHTGRSGG- 1473
QY	1538	PPAFTSTNVSAALALNRTSQSISILKRAEPVTSNDISELTSTL-----CKHFK 1587
Db	1474	---GNKLVTIDKAAQSHHETAQSTTFEGKQVVLQAGNDANILGSNVISDNGTIQAGNHVR 1530
QY	1588	DSATTWLAALKELDDAKPAEQHLIQOHFSKDVVDGDEREAVRNLKLVIRQQAADSH 1647
Db	1531	IGTT-----OTQSOSEYVHTQKSLGMSAGIG-----FTI-----GSKIN 1565
QY	1648	SMELGSASHSTTVNNLSRINNNDIIVELLHKHDAALPASSAKRELGEMM-----NNDPAK 1702
Db	1566	TQENQSQSNEHTGSTVSLKGDTTI-VAGKHYE-----QIGSTVSSPEGNNTIYAQ 1615
QY	1703	DIIKQLQSTPFSASVSMELKQGLREQTEKAILDGKVGREVCV-----LFQDRNLRV 1756



Db 1616 SIDIOAAHKLNSNTQTYEQKGLTVARSPVTD--LAQQAIAVAQSSKQVQSKND-RV 1672  
QY 1757 KSVSVSQS--VSKSGFNTPALLGTSNSAAMSERNITGTFKYQCDQNTPRRT-LEG 1813  
Db 1673 NAMAANAAGWQAYQTKGAQNLANGTNAKQVS-----ISITVGEQCN--RQITQVQA 1723  
QY 1814 GIAQANPOVASALTDLKKEGLEMK 1838  
Db 1724 NQAQASQIQAGGKTTLIATGAEQS 1748  
  
RESULT 5  
A35938  
Profilaggrin - human (fragments)  
C:Species: Homo sapiens (man)  
C:Date: 14-Dec-1990 #sequence\_revision 02-Jul-1996 #text\_change 29-Sep-1999  
C:Accession: A35938  
R:Gan, S.Q.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.  
Biochemistry 29, 9432-9440, 1990  
A:Title: Organization, structure, and polymorphisms of the human profilaggrin gene.  
A:Reference number: A35938; PMID:91064347; PMID:2248957  
A:Accession: A35938  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-2248 <GAN>  
A:Cross-references: GS:J02929  
C:Genetics:  
A:Gene: GDB:FLG  
A:Cross-references: GDB:I19912; OMIM:135940  
A:Map position: lq21-lq21  
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology  
C:Keywords: Ep hand; epidermis; polymorphism; tandem repeat  
F:246-565/Region: profilaggrin repeat  
F:570-893/Region: profilaggrin repeat  
F:1074-1397/Region: profilaggrin repeat  
F:1573-1896/Region: profilaggrin repeat  
  
Query Match 2.7%; Score 258.5; DB 2; Length 2248;  
Best Local Similarity 18.4%; Pred. No. 0.00024;  
Matches 381; Conservative 249; Mismatches 815; Indels 629; Gaps 83;  
  
QY 7 GTEHKAHVTAHNPVGHVQALQ-----QGSSSSSPQNAASLAAGKMGKMPR 56  
Db 487 GSGHQASDSRHSIGHGQASSAVRDSGHRGYSQASQOEGHSENSQSVSGQAR 546  
QY 57 IHQPTAADGISAAHQKKSFLRGLGKTKKFSRQAPQGPQGTTHSKGATLRDLLARDG 116  
Db 547 SHQQS-----HQES-----TRGQSRGSRGSR--GSFLYQVSTHQ 580  
QY 117 ETOHEAAAPAAALTRSGGVKRRNMDMAGRPVYKGGSGEDKVPQOKRHQLNFGQMRQ 176  
Db 581 ESAHRSAPSTR-----RQSHHDQA-RDSSRHSASQEG--D 616  
QY 177 TMLSKMAHPASANAG-----DRLQSPHPIGSHS-----ELKEPVGSTKAT 221  
Db 617 TI---RGHPSSRGQSGHYEVSVDGSHS-----GSHSHTTSQGRSDASHTGSR 668  
QY 222 AHADRVIAQEDDDSPFQQLHQQLARENPQPKLGVATPISARFQPKLTAVAESVL 281  
Db 669 ASRQTRDEQSGDGRHSGSHHQAESTQADS-SRHQVGQGGAGSR-----TSRN 718  
QY 282 EGTDTTQ---SPLKPSMLKSGAGVT-----PLAVTLDKGLQAPDN 322  
Db 719 QGSSVSQDRDEQSGEDSERHSGSASNRHSGSAEQSRDGRPHGSHDEDRAGHQADS 778  
QY 323 PPAINTLLKQTLCK-----DTQVLAHHAGSDGSHLLLDNKGHLFDIKST 368  
Db 779 SSQSGTRHTQTSSRQAASSQEQARAGDRHSGSHQQAADSRH-----SGIGRG 829  
QY 369 ATSYSVLHNSHPGEIKGLAQAGTGVSVVDGSGKTS-----LGSQTQSHNKTMLSQGEA 424  
Db 830 QASSAVRDRGHRGS-RGSAQSQEGH-SEDSQSQSVSAQRAQSHQSQHQUEST----- 880

QY 425 HRSLLTGTIWOHPAGAARPOGES-----IRLHDDKIHLHPGLGWQSAKDQTHSQL 475  
Db 881 -----RGRSQSGRSGSYQVSTHQES-----ESAGRTTST 915  
QY 476 SRQADGKLYALKDNRTLQNLSDNKSSEKLVKIKSYSDQDQGVAILTDTQGRHKSIMP 535  
Db 916 GRRQGSHEQARDS-----SRHSASQSGQDTIRAHPCGSRG-----GRQSHHQ 960  
QY 536 SLDASPEHSILSLHFAHAGHLLHKGSELAEQAQSVASHGRILVADSGKLFSAALPKQG 595  
Db 961 SVDRS-----GHSGSHSHTTSQGRSDA-SHG-----QSG 989  
QY 596 DGNELKMKAMPQHALDEHFGHDHQSFFHD--DHGQLNALVKNNFRQHQHACPLGNDHQF 653  
Db 990 SGSH-----HQGADSRHSIGHGQASSAVRDSGHRGSGSGQASDSEG 1033  
QY 654 HPGWNLTDALVIDNLQGLHTNPEPHEIDMLDGLSLALQEGKLY-----FDQ----- 702  
Db 1034 HSEDSDTQSVSAQAGPHQ---QSHQESTRGSAGRSRSGSFLYQVSTHQESAHGR 1090  
QY 703 ---LTKGWTGA-----ESDCKQLKGLDGAAYLLKDGVEVKRLIN----- 739  
Db 1091 ARTSTRGQSHHQAQDSSRHSTSQEGQDTIRHPGSPSGRHSYEQSVNSTGSHGS 1150  
QY 740 -----QSTSSIKHGTENVFSLP-HVRNKPEPGDALQGLNK-----DDKAQANAV 782  
Db 1151 HHSHTTSQGRSDASHGTSRGSASRETHNEQSGDGRHSRHSRHSQASWADSSGSHQAG 1210  
QY 783 IGWNYLALTEKDIRSQIKPTQQLERPAQTLSREGISGELKDIIH--VDHKQNYALT 840  
Db 1211 QGSSG-SRTSRNCGSSF-----SQSDSQSGSEDSERRSGSASRNRHSGSAREQSRDGR 1264  
QY 841 HEGEVFHPREAWQNGAESWHKALPQESKLSLDMSEHHPDIATFEDGSOHQLKA- 899  
Db 1265 HPGS-HHEDRAGHGSADSRQSGTRHTQTSSRQAASSQEQARSGRHSYEQSVNSTGSHGS 1323  
QY 900 GGHAYAAAPBGRPLAV-----GTSGSQTVFNR-----LMQGVKGVIPGS-----GL 941  
Db 1324 SSRHSIGRQASAVRDSGHRGSRGQASDNEGHSSEDSQSVAGORAGSHHSHQES 1383  
QY 942 TVKLQATQGTGTAEGKVKSKFSERIRAYAFNPTMTSPRPIKNAAYATQHWGQREGKL 1001  
Db 1384 TRQSRSTGRSGSFLYQVST-----HEQESSHGWSGTS--- 1418  
QY 1002 PLYEMGALIKOLDANVRHNAQPDLOSKLETILD-----GEHGAELNDMKRPRDELE 1056  
Db 1419 -TRGQSGHHEQAQ-DSSRHSASQ-----EGQDTIHGHPGHS-----DSSR 1459  
QY 1057 QSATRSVTILGQHGQVGLKNGEINSEFKPSPGKALVQSFVNRS---GQDLKSL----- 1108  
Db 1460 QSGTRHTESSRQAA--SSHEQARSSAGERHSHHQQSADSRHAGHGHQASAVRDSG 1518  
QY 1109 -----QAVHATPPSABSKLQSMGLHFVSAGVDMSHQKEIPLGRQRPNDKTAITKSR 1162  
Db 1519 HRCYRQSOATDSEGHSESDTSQSVSAQAGPHQAQ--ESARGQSGESSG-----RSG 1571  
QY 1163 LILDTVTIGELHELADKALVS-----DHKPDADQIKQLRQFDTILREKRYENPV 1213  
Db 1572 SFLYQVSTHQESTHGQSVFSTGGQGGHDDQAQDSRHSASQEQDTIRHPGSSR-- 1629  
QY 1214 KHYTDMGFTNKALEANYDAVKAFINAFKKEHGVNLTTRTVLESQGSALAKKNTLL 1273  
Db 1630 -----GRRHSHYEQSVDR-----SGHSGSHSHTTSQGRSDASHGSGSRASQTRN 1678  
QY 1274 SLDGSEMSFSRSY-----GGSVTVFPTLSKKVPVPIPGAGITLD- 1316  
Db 1679 EQSGDGRSHSHHHEASTQAESRHSQAGQSA--GPRTSR-----QGSSVSQDS 1730  
QY 1317 -----RAYNLSPRSTGGLNVSGRGGVSGNIMVATGHDVMPYMTGKTSAGNAS 1367  
Db 1731 DSQGHSEDSERSGSASNRHSGSAQEQSRDGS-----RPTSHH--EDRAGHRHSA-ESS 1782





QY 1800 GQDQWTPR-----RFTL-EGGIAQNPQVASALTDLKKEGLE 1835  
 Db 3231 GGDQHAQQLGSDLGSVIADVGFALAAAGTAKAAEILGEAGISLSKDYLE 3279  
  
 RESULT 7  
 E82589  
 hemagglutinin-like secreted protein XF2196 [imported] - Xylella fastidiosa (strain 9a5c)  
 C:Species: Xylella fastidiosa  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: E82589  
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: E82589  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3442 <SIM>  
 A:Cross-references: GB:AE004032; GB:AE003849; NID:99107324; PIDN:AAF84995.1; GSPDB:GN00  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier,  
 as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.  
 Submitted to Genbank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Froh  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.B.; Kuramae, E.E.; Laig  
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins,  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; ;  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF2196  
  
 Query Match 2.6%; Score 247; DB 2; Length 3442;  
 Best Local Similarity 18.7%; Pred. No. 0.0018;  
 Matches 414; Conservative 292; Mismatches 821; Indels 692; Gaps 100;  
  
 QY 3 LKSLGTEHKAHVTAHNPVGHGVALQGGSSSPQNAASLAAGKRNKMPRIHQPST 62  
 Db 1374 IDNLGTGRLYGDHIAL-----HAQTLNTRDSTGDGHTHAATAARQLDIGADTLRNTAN 1428  
  
 QY 63 A---ADGISAAHQKKSFSLRGCL---GTYKF--SRSAPOGPGGTHSKGATLRDLLARD 114  
 Db 1429 AMILSGDDAAI-----GATLDNALHATGTATLDNRSATDITGTLLNTITTTLLNNI--RD 1481  
  
 QY 115 DGETQTEAAAPDA---ARLTRSGGVKRRNMDMAGRPWVGKGGGDEKVPQCKHQHQLNNF 171  
 Db 1482 N---VHIAHAPDVVTEARMEQPHW--RKQPN-----GGSGNPFSTNYDAHD;--- 1525  
  
 QY 172 GQMRQTMLSKMAHPASANAGDRLOHSPPHIPGSHHEIKEPVGSTK-----A 219  
 Db 1526 -----YILNPADIKKD-----PVITPDGQIHRAIVRLTPTQTSAYFYARGGLYA 1570  
  
 QY 220 TTAHADRVELAQEDDDSEFQQLHQQLARERNPPOPPKLGIVATPTISARFPQKLTAVES 279  
 Db 1571 SQAERRRMDLTARTGDSVVLYYTDR-----ODKQPNPDHVA---ATNHS 1613  
  
 QY 280 VLEGTTTQSPKPSQMLKSGAGVTPFLAVTLDKGLQAPDNPALNTLLKQLGKDTQ 339  
 Db 1614 AFIGLDTPOONERFQI-----VFITVAPGDDRLTYDSNYG-----TCTDDCV 1655  
  
 QY 340 HYLAAHSSDGSQHLLLDNKGHLFDIKSTATSYVLNHSHPGIEIKGLAQAGTGSVSDG 399  
 Db 1656 RLVTWHDYTD-PDHTLID-----MHRG-PNDVR-----DN 1683  
 QY 400 KSGKTSIGSGTOSHNNKTMV-SQGRAPHSITLTGTHOHCACAPDQCESTH----- 451

Db 1684 EKYR-----DATRTTQDILNPDAPALIQTG-----GAMMIQTDTLRNHYADLLAGGD 1733  
QY 452 DKHIIHPELGVQWQADKTHSOLSRQADCKLYALKDNRTQLNLSKSEKLVDKIKSY 511  
Db 1734 QTVIGLPHPTKESDDEHKY-----KRVLLIDNRLAQ-LS-----RTDTFHHI 1776  
QY 512 SVDPQGOVA-----LITDTPEGRHKMSIMPSLDASPESHISLSLHFADAHQGLLHG 561  
Db 1777 STTYRGKVPWNESESTTTTQIGGRITSGGHCHIAAQTLNNVDTSHAPEPIQLHTYN 1836  
QY 562 KSELEAQSVASHGRVLVADSEKLFSAAIKPKQGDGNELKMKAMPQHALDEHFHGHQIS 621  
Db 1837 PS---TQTLVWNGVITVDTSPSLHTVSLADNGSPSAQELTYIPDQSIITP-----1885  
QY 622 GFPHDDHGQNALVK-----NN-----NQOHACPLGNDHQFHPGMN 658  
Db 1886 -----NAPTRDPAAPAVTVPTGFLTLPNNSLFTLHPDAATLITDPRFTTGRP 1935  
QY 659 LTDALVIDNQLGHHHTNPPEHIELDMGLHLSALQEGKLYHFDOLT--KGWTAESDCQ 716  
Db 1936 YTSADTQHLALGDHDT--LHKRLGDGYEQRLIRE---QLAQLTGRRRLDGYTDDDDQ 1988  
QY 717 LKXGLDGAAYLLKDGVEVK-----RLMINQSTSI-----KHGTENVFSLPHVRN 760  
Db 1989 YRALLDAGVTVAQHQLRPGIALSADQLAQLTSDIVMLVQDVLPGDPTTTRALVPRLYL 2048  
QY 761 KPPEGDALQGNKDKAQAMAVIGNVKYLALTEKGDIRSFOIKPGTQOLERPAOTLSREG 820  
Db 2049 RPRTGD---LTPDGAALLAASTTINAH-TFTNTGTTIDARHL-----IDINAHMDOQG 2097  
QY 821 -----ISGELKDIIHVDHKQNYALPH-EGEVFHQP--REAWONGAESSSWHKL- 865  
Db 2098 GRLTADAIHHTTGDFITLGGQFARGYLKVHAQGNFLASSTLRDATTTQGRHHSVTELD 2157  
QY 866 -----ALPOSEKLSLDMSHHKPIATEDSGSQHOLKAGWHYAAPERGLAVGT- 917  
Db 2158 QOAGFTVTGPGAYLGL-STDQAMTHNGVAIINNITGD-----GYTSINA--TGFLHLGTL 2208  
QY 918 -----SGSOTVFNRLMQGVKVIKPGSLTVKLSA-----947  
Db 2209 TTHRSDDTQMDPRNSHRIDTEYGTSTIG-NGDIQLNSGQDINLRATLHSTQGTITAL 2267  
QY 948 QTGGMT-----GAEGRKVSXKFSERIRAYAPNPTMSTPRPIKNAAY 988  
Db 2268 ATGNVTIHTGDTIQYTSQDSHTKSGSLNSRTTTHAQOQQTQALGSTLSADKVFVKGN 2327  
QY 989 ATQHWQOQREGKPLYLEMQGALIKOLDANVR-----HNAPODLOSQKLETLDLGHHGAE 1044  
Db 2328 ITVTGSHVVSDAGTYMQAEHDLTLQAATHTTQSTYSHTKQRL-----IRNGGASL 2379  
QY 1045 -LNDMKRFRDEQSATRSVTVLGOHQG-----VLKXNGEINSEFKPSP 1087  
Db 2380 TLGQSQSDTSTTTATTTTGLIGATNGNVLLAGHYQQIGSDVLSLPHGDIDIIHAKVD 2439  
QY 1088 GKALVQSNVNRSGQDL-----SKSLQQAHVATPSPAESKLSQMLGH 1129  
Db 2440 ---IIQAHTSQTQHTATROSLTVALSTPLIAGTAQQAQQAARSGDPRLQALAGL 2496  
QY 1130 FVSAG-----VMSHQ-----KGEIPLGRDNDKNTALTKSLIILDTVTIGELHELA 1177  
Db 2497 TTALGAKNTIDAVRQDPRALGGLNASLTVGRSTHSDTTTTTTSTAAGSNVTAGGNVHISA 2556  
QY 1178 D-----KAKLVSDHKPDAD-QIKQLRQOPDTLREKRYENPV-----1213  
Db 2557 TGDGTASTLTIQGSVDVRGDTMYLKADGDIALAAQNTVNTQDRNGRSAGVGVAVNLGS 2616  
QY 1214 -----KHYTDMGFTHNKALEANYDAVKAFINAFKKEHHGVNLTTRTVL 1256  
Db 2617 GTSAGLTAHASTTSGSGHSTDLTWSNVHVGGNLLAIDAGDGLMK---GAIGTAKHVI 2673  
QY 1257 ESQGSAAELAKLNTLLSDSGESMSFSRSYGGGVSTVFVPTLSKKVPVFPVPGAITLUD 1316  
Db 2674 ----ADIAGNLRIESLQ-DTHQYRSTDRSLGSLTA-----GAGFSGS 2711

QY 1317 RAYN-----LSFRTSGGLNVSFGRDGVSGNINMVAIGHDVPVMTGKTSAGNASD 1368  
Db 2712 ANLNHQITIRSDYASVTEQSG---LFTGDGGYQLTVGGQT-HLIGGAITSNIAHGNLN 2766  
QY 1369 WLSAKHIKISPLRIGRAVSGTL-QGTILQNSLXFKLIEDELPGFI--HGLTHGTLTPAELL 1425  
Db 2767 SLD-----TGTILQNIENHANTATQVNLGGYSGRNGTGT-----2804  
QY 1426 QKGIHQMKQSGKLTFSVDTSANLDRAGINLNEDSGKPNGVTVARVS-----AGLSASA 1479  
Db 2805 -----DQGHATATQVEGTLTPSHNGLS---ASPPSAMTARDSSHSTYSIGISOGA 2853  
QY 1480 NLAAGSRERTTSGQFGSTTSASNNRP-----1506  
Db 2854 LTRDDDTAQHALTGHATAETIATILNRDLTDTATSNALTPIFDBQRINAGFDIVSGLQRE 2913  
QY 1507 --TFLNVGVGAGANL-----TAALGVAH-----SSTHEGKEV--GIF 1538  
Db 2914 TGTFINNRAEADLTKRQATAADHAAHDPSNGFNDQORHARERQAIATNEAHAIKDAMG 2973  
QY 1539 PAFSTTVNSAALADNRTSOSISLELKRABPVTNSDISELTSTLGHKF-----KDSAT 1591  
Db 2974 PGTYRQITTTALAAG--ASGNVS-----AASNVSAASSDLAKHVIWVYVQOOGATA 3023  
QY 1592 TKMLAALKELDDAKPAR-QLHIL-----QQHESA-----KD 1621  
Db 3024 IGHVATGQLTGEGSPLHAALHALLACAGAAASQOHCSSGAGAAASSVLTLGFSDDRPED 3083  
QY 1622 VUGDERYEAVRNKLIVIROQAADSHWELSGSASHSTYNNLSRINNNGI-----VELLH 1676  
Db 3084 TTQDR--EAKRNLITSTVTIAGSTGNT-DAATAHAA-----IAAVDNNWLAQKQVQVMIN 3136  
QY 1677 KHPDAALPASSAKRLGEMMNDPALKDIIKQLQSTPFSASVSMELKDGLREQTEKAILD 1736  
Db 3137 EEFERA-----TEKKGRL-----EEKVRAKWRB-IDARQDKLTVD 3172  
QY 1737 GKV-GREEVGVLFDORNNLR---VKSVSVSVOSVSKSEGFNTPALLLGTSNAAAMSERNI 1792  
Db 3173 GLLAKGLKESGI--NDINGLEHLFHPVDTVHELKGI--LTHPTLLQLGESAVOELLNKV 3228  
QY 1793 GTIN--FKYGOQNTPRFRTLEGG--TAQANPQVASALT-----DLKKEGLE 1835  
Db 3229 SRMEALLVGGDOHA-QQFGEDLGSVIADVGVALLAAAGTFKAAETLGEAGINLSKDVLE 3286

## RESULT 8

H81193  
hemagglutinin/hemolysin-related protein NMB0493 [imported] - Neisseria meningitidis (str  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: H81193  
R:Testelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC59.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: H81193  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2703 <RTE>  
A:Cross-references: GB:A8002405; GB:A8002098; NID:g7225708; PIDN:AAF40927.1; PID:g7225711  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB0493

Query Match 2.6%; Score 245; DB 2; Length 2703;

Best Local Similarity 19.0%; Pred. No. 0.0015;

Matches 414; Conservative 282; Mismatches 766; Indels 712; Gaps 105;

QY 10 HKAAVHTAAHNPVGHGVALQQGSSSSSPQNAASLAIEGKNRGMKPRIHQPSTAAD---G 66





Db 2080 FKVLDDTQGEARVFLADNLKLKKELOSNKESVSOMKQD-----EDLERRLEQAEEKHL 2135  
QY 1587 KDSATTWMLAALKELDPAEQHILHQQHFSADKVVVDREYAVRNKLKLVIRQQAADS 1646  
Db 2136 KEKN-----MQBKDALRERKHL-----BETIGEIVTUNKDKQKQVQLOENLDS 2182  
QY 1647 HSMELG--SASHSTYNNLRINNDGIVELLHKEHFDALPASPASSAKRILGEMNNDPALKDI 1704  
Db 2183 TVTQLAAPTKSMSSLODDRVR-IDEAKWKERKFSDAIQSKEEBIRLKE--DNCSVLKQD 2239  
QY 1705 IKQLOSTPFSASVSM-ELKDLG-REQTEKAILDKGVRE----- 1742  
Db 2240 LRQM-----SIHMBELKINISRLRHDQIWESKAQTEVQLQCKVCDTLQGENKELLS 2291  
QY 1743 -----EVGVLFODRNLNRVKSVSQSVSKSEGFNTPALLLGTSNSAASM 1788  
Db 2292 QLETRHLYHSSQNELAKLESLKSLKQDLTDLNSLEKCKEQK-----GNLEGIRQ 2345  
QY 1789 ERNIGTINFKYGQ---DONTFRPT--LEGGIAQANPQVASALTDLKKEGLEM 1836  
Db 2346 EADTONSKSYEQLETDLOASRELTLRLHBEINNKKEQKIISLLSG-KEEAIQV 2397  
  
RESULT 10  
I52300  
gi|152300|human  
N;Alternate names: gcp372  
C;Species: Homo sapiens (man)  
C;Date: 01-Nov-1996 #sequence\_revision 01-Nov-1996 #text\_change 20-Aug-1999  
C;Accession: I52300  
R;Sonda, M.; Misumi, Y.; Fujiwara, T.; Nishioka, M.; Ikehara, Y.  
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994  
A;Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in  
A;Reference number: I52300; MUID:95100974; PMID:7802676  
A;Accession: I52300  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-3225 <RES>  
C;Cross-references: GB:D25542; NID:G662389; PIDN:BAA05025.1; PID:G808869  
C;Superfamily: giantin

Query Match 2.6%; Score 243.5; DB 2; Length 3225;  
Best Local Similarity 18.6%; Pred. No. 0.0024;  
Matches 389; Conservative 311; Mismatches 816; Indels 577; Gaps 92;  
  
QY 23 GHGVALQOQSSSSPQNAASLAAGKNGKMPRIHQ--PSTAADGISAHQKKSFLR 80  
Db 569 GEGIA-----PIKMKVLEDTGQDFPLMNESSILPAVEKQASTBHQSTSEHI- 618  
QY 81 GCLGTKKFSRAPQGFTHSGATLRDLLARDGGETQHEAAPDAAR----- 129  
Db 619 -----SLNDAGVELKS--TKQDGD-XSLSAVPDQGCHQDELERLKS 657  
QY 130 -----LTRSGVKRRNMDMAGRPVKGSGEDKVPQQRHQLNN-----FGQMR 175  
Db 658 QILELELNPHKQEIYEKNLDEKA-----KEISNLNQLIEEFKKNADNNSAFTALSEER 712  
QY 176 QTWLSKM-----AHPASANAGDRLOHSPPHIPGSHHETKEBPVGSTSK 218  
Db 713 DQLLSQVKELSMVELRAVQVQLXNLAERQRRLDYES--QTAHDNLLTQIHLSLI 769  
QY 219 ATTAHADRVRIQED--DSEFQQLHQORLARENPFPQPKGV-----ATPISARFP 271  
Db 770 EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRLSQSQLOKNESEVLGAEVRVHISGVVE 829  
QY 272 KLTAVAESVLEGGTDTQSPKLP-----QSMKSGSAGVTPPLAVILDKKLQALPDN-- 322  
Db 830 LSQALSQKELEITKMDQLLEKKRDVETLQOTIEEKDQVQTEISFSTMRKQVNLNEEKS 889  
QY 323 -----PPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKGHLFDIKSTATSYV 374  
Db 890 LGVEIKTLQELNLLSRAEAKKEQVEEDNEVSSGLKQ-----YDEMSPAGQISK 940

QY 375 LHNSHPGEI-----KGLAQAGTGSVSDGSGKIS---LGSQTSHNKTMLSOP-- 421  
Db 941 EEOHEFDLLKXENQKRLQOALINRKLQVRSLREBELANLKESKKEIPLSETER 1000  
QY 422 GEARSHLLTGIWHPAGAARPOGESIRLHDDKIHILHPELGVMSQADKTHDSQLSROADG 481  
Db 1001 GEVEEDKENKEYSEKCVTSKQEIYI---KOTISEKVEL-QHIRKDLKLEKLAEEQF 1056  
QY 482 KLYALKDNRTLN-----LSDNKSEKLVDKIKSYVD--ORQGVAILTDTFGR 528  
Db 1057 QALVKQNNQTLQDKTNQIDLLQAEISENQA---IIQKLITSNTDASDGSVALVKET--- 1110  
QY 529 HKMSIMESLDSPESHISLSLHFADAHQGLHKGSELEAQSVAISHGLRVVADSEGLKFS 588  
Db 1111 --VVISPPCTGSE-----HWKPELEKLLALEKEX---EQLOKKLOE 1148  
QY 589 A-----AIPKGD-----GNELKMKAMPOHALDEHFGHDHQSISGFFHDHGOALNALVKN 637  
Db 1149 ALTSRKAILKKAQEKERHLREELKQKDDYNRLQEQDEQSKENENIGDQLRQLQIQVRE 1208  
QY 638 NF-----RQOHACPLGNDHQHPGWNLTALVIDNQLGLHHTNPEPHEILDMHGLSL 690  
Db 1209 SIDGKLPSTQOESC-----SSTPG--LEELPKATE--QHTTOP-----VLE----- 1247  
QY 691 ALQEGKLHYFDQLTKGWTGAESDCKQLKGLDGAAY--LLKDGVEVKR---LNINQSTSS 744  
Db 1248 -----SNLCPDWPSHSEDSALOGGTSVAQIKQAQLKEIAEKVELELKVSTTSE 1297  
QY 745 IKGTENVFSLPHVRNKPPEGDALQGLNKDDKAQAMAVIGVNYKALATEKGDTRSFQIKP 804  
Db 1298 LTKKSEVFLQEQINK-----QGL-----EIES--LKT 1324  
QY 805 GTQQLERPAOTLREGISGELKDIHVHDKQNLALYALTHEGEVFPQPRE--AWQNGAESSS 861  
Db 1325 VSHEAEVHAESLQOKLESSQLIAGLEHLRELQPKLDELQKLSKKEEDVSYLSGQJSEK 1384  
QY 862 WHKLA-----LPOSESKLSLDSHSHKPIATPEDSGHQKAGGHHAYAAPRGPLAVG 916  
Db 1385 EAAALTKIQTETIEQEDLIKALHTQLEWQ--AKEHDERIKQLQVELCEMKQKPEE---IG 1438  
QY 917 TSGSQTVFNLMOGVKGVIPGSLIVKLSAQTCGTMGAERKVSXKFSERI---RAYAF 973  
Db 1439 EE-----SRAKQOIORKL-----QAALISRKALKENKSLQBELSLARGTIE 1480  
QY 974 NPTMTSTPRIPKNAAYATQHCWQGRE-----GLKPLYEMOGALTIKQLDHNVHNAPOPD 1027  
Db 1481 RLTKSL-----ADVESQVSAQNKEDTVLGRLLALQEEERDKLITEMD-----RSLLENQS 1530  
QY 1028 LQSKLETDLGEGHABELLNDMKFRDELEQASRVSIVLQHQHGVLSKNGSEINSEFKPSP 1087  
Db 1531 LSSSCBSLKALEG--LTEDKELVKIEISLKSXKIAESTEWO--EKHKELQKEY---- 1581  
QY 1088 GKALVQSFVNRSGQDLKSLOQAVHATPPSAESKLSQMSLGHFVSGVADMVSH-----QKG 1142  
Db 1582 -EILLOSQ-----ENVSNAAEIRQHVAVRQEK-QELYGKLRSSTANKKETEKOQLEA 1633  
QY 1143 EIPLGRQDPNDKLTALTKSLILDTV-----TIGELHELADKAK---LVSPHKPDADQ 1192  
Db 1634 EQEMEEMKMKRFAKSKQKQKILEEENDRLRAEVHPAGDTAKECMTWELLSNASMKEE 1693  
QY 1193 IKQLRQOQFDTLRK-----RYESNPVKHYTDMGFTHNKALEANYADVAKFINAF 1241  
Db 1694 LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE---GNVSKQANLEAT----- 1742  
QY 1242 KKEHHGVNLTTRVLESQSAELAKKLKNTILSLDS-----GESMSFSRSGGVSTVF-- 1295  
Db 1743 EXHDNQTNTVEGTQSIGETE-----EQDSLSMSTRPTCSFSPVSAKSNPAVSKDFSS 1797  
QY 1296 -----VPTLSKKVPVPIPGAGITLDRAYNLSFSRTSGG-----LNVSFGRDGV 1340  
Db 1798 HDEINNYLQOIDQIKERI-----AGLEERKQKNEFQSTLENEKNTLLSQISTKQDEL 1850

1341 QY -----SGNIMVATGHVMPYMTGKTSAGNASDWLSAKHKISPLDRIGAAVSGTLOG 1392  
1851 Db KMLQEEVTKMNLNQIQBELSRVTKLKEAABEKD-----DLEE 1890  
1393 QY TLONSILKFKLUTEDBLPGFIHGLTHGILTPAELLQKGLHOMKQSKLTFVSVD----- 1444  
1891 Db RLMNQLA-----ELNGSI-GNYCQDVDTAQIKNELLESEMKNLKCVCSELEEKQOLVK 1943  
1445 QY --TSANLDDR-----AGINLNEGSKPNGV-----TARVSAGLS 1476  
1944 Db EKTQVSEIRKEVELEKIQGAQ-KEPGNKHAKELQELLKEKQEVKQLQKDCIRYQEKIS 2002  
1477 QY ASANLAAGSRERSTSQFGSTTSASNNRPTFLNGVGAGANLTAALGVAHSTHEGKPVG 1536  
2003 Db -----LERTVKALEFVQTESQKDLITKEN-----LAQAVEHRKKAQAEALS 2045  
1537 QY IFPAFTSTNVSALAL-DN-----RTSQSISLEIKRAEPTVTSNDIS-ELSTLTKHF 1586  
2046 Db FKVLDDTQSEAAARVLADNLKLLKELQSNKESVKSQMKQD-----EDLERRLEQAEKHL 2101  
1587 QY KDSATTKMLAALKELDPAEQHLILQHFSAKDVVGDERYEAVRNKLLKLVIRQQAADS 1646  
2102 Db MOEKLDALREKVL-----BETIGEIVTLNKDKQEVQQLQENLDS 2148  
1647 QY HSMELG--SASHSTYNNLRINNDGIVELLHKHFDALPASSAKRILGEMMNDPALKDI 1704  
2149 Db TVTQLAAFTKSMSSLODDRVR-IDBAKKWERKFSDAIQSKEEBIRLKE--DNCSVLKQD 2205  
1705 QY IKLOSTPFSASVSM-ELKXGL-REOTEKAILDKVGRE----- 1742  
2206 Db LRQW-----SHMEELKINISRLBHDQIWESKAQTEVLOQKQVCDTLQGENKELLS 2257  
1743 QY -----EVGLFODRNLNRVKSVSQSVSKSEGFNTPALLLGTSNSAAMS 1788  
2258 Db QLETRHLVHSSQNELAKLSLQDLTDLNSLEKCKEQK-----GNLEGIIRQ 2311  
1789 QY ERNIGTINFKYQ---DQNTPRRPT--LEGGIAQANFQVASALTDLKKEGLEM 1836  
2312 Db EADIQNSKFSYEQLTDLQASRELTSRLHBEINNKQKIIISLLSG-KEEAQV 2363  
RESULT 11  
AC0304  
probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 02-Nov-2001  
C;Accession: AC0304  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AC0304  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-2535 <R>  
A;Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175  
C;Genetics:  
A;Gene: YPO2490

Query Match 2.6%; Score 241; DB 2; Length 2535;  
Best Local Similarity 19.4%; Pred. No. 0.0022;  
Matches 386; Conservative 232; Mismatches 659; Indels 710; Gaps 94;  
QY 265 ISARFPQKLTAESVLEGTDTTQSPKPSQMLKSGAGVTPLATVLDKGLQ-----L 318  
Db 416 LSVEASGKITAGKNI---SGTAQRTAHLRLDLSGQTSQRDITLIAQGEIDLTAELL 472  
QY 319 APDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLILDNKGHLFDIKSTATSVLHNS 378  
Db 473 ASDRLSAATTALLRT---DNASLIAEQITLDAQ---ALSNVGG---IAHTGTTFDNL--N 522

379 QY HPGEIK---GKLAQAGTGSVSD--GKSGKISLGSQTQSHNKTMLSQPGEAHRSLLTGIV 433  
523 Db LPGDVNRGTTLLSSGTLSLQABSLNSNGSLGAGVQSDGR--LTEIGDLRVTTROQLI 580  
434 QY OH-----PAGAARQGESIRLHD-----DKIHI----- 456  
581 Db AHGQTLAAGAMALTGRIDLADSYTOAREMTLTANRGDISTQBATVLAULTLSINTAQL 640  
457 QY -----LHPELGVW-----QSADKDTHSQLSRQADGKLYALKNRTIQLN 495  
641 Db NNQGGTLAGNTLALDQFPDQCGQVTSQDLITDLQRDFSHQAGSTLOAGRDLTLTSL 699  
496 QY SDNKSSEKLVDKIKSVSVQORQVAILUTDPGRHKMSIMPDLSPASHISLSHFADAH 555  
700 Db G-----AVTNDGQL-VAGGTLSTHSDSLNS----- 724  
556 QY QGLHKGSELEAQSVAISHGRVAV-----ADSE-----GKLFSAALPKQDGNELKMKAM 605  
725 Db GNLIATQAEUNATGALINHGEIITLGGLODTSNTLFWTGSIIISA-----EATLNAR 775  
606 QY PQHALDEHFHGHQISGFFHDDHQLNALVKNFRQOHACPLGNHGHQHPWNNLTDLVI 665  
776 Db ERIT---NSGPDALIGA--TDENGLTALL-----APVI 803  
666 QY DNQGLHHTNPEP-HEILDM-----GHLGSLA-----LQEGK--LHVPDQ 702  
804 Db ENSDTVHTTDTAPTITTLGMTVILAGGHARDGHYASAAQVNLNLSGIESKMDLIYATT 863  
703 QY LTRG-----WTGAESD-----CKQLKGLDGAAYLLKDGVEKR 735  
864 Db LTRSRHLLTANTDPIVADTVTGTAVTAENPDIPGRYAEPPDGAANSYI----- 915  
736 QY LNIQSTSIKHGTENVFSLPHVRNKPDPGDLQGLNK-DDKAQAMAVIGNKYLAIT-E 793  
916 Db -----GTEYTSVI-----AYNGIDQISPEAQLLA--CGN-----LTPQ 946  
794 QY KGDIRSFOIKPGTQ-QLERPAQTLRSREGISGELKDI-----HVDHKQNLVALTHE 842  
947 Db VGTLENFWSKVSQAQGEIDLTGVTLOQDQGWGDQORLMEQTTSQSSVWRYRTYKGLWTRBW 1006  
843 QY GEVPHQPREAW-----ONGAESSSWHKLALPQSESKLSLSDMSHEHKPI 886  
1007 Db PEVSEATSEYASFTAKTLSSGTTINNGANPGA-----TAPPADRNSGKDLAVEFNGI 1062  
887 QY A-----TFEDGSQHQLKAGHAYAAPE-----RGP---LAVGTSGSQTVFNRL-- 927  
1063 Db SLTQPNGLYQFTDHTVGGGGLIETHAPANLNWRGSDYVLQQLNNDPDVIFKRLGD 1122  
928 QY -----MQGVKGKVPESGLTVKLSAQOTGTMGAERKVSSEKESIRA-----YAFNPTMS 978  
1123 Db NAYEQRLVRDQVL-----ALTQAVASDYRSQAQEQFEALFAAGLEYSKAFNALG 1172  
979 QY TPRPIKNAAYATQH-----GWQREGELKPLYEMQGALIKQLDAHNVHNAPODQLQSK 1031  
1173 Db THLSAEQMAALTHNIVLMETRDVAGQTVLVPVVVLVAGVKPGDGLQANGALIAENISL 1229  
1032 QY LETLDLGEHGA-ELLNDMKFRDEQSATRSVTVLGQHGVKLSNGE-----INSEFKP 1085  
1230 Db TEVOGFTNAGAITATNDLK-----ISMAQDIT-LNNRGLLQAGGDMQLSTLNSIDL 1281  
1086 QY SPGKALVQSFNVNRSGDLSKLOQAVHATPPSAESKLQSMGLHF-----VSAGV 1135  
1282 Db TSARINATNLQLD-SGRDVLRLTDSQAQLSSDNGAVSRDQTLTGPLASINVSNNAINTGR 1340  
1136 QY DMSHOKGEIPLGRORDPNDKTLTKSRLILDTVTIGELHELADKAKLVSDHKPDADQIKQ 1195  
1341 Db DFIMQASLNVGQ-----DLQVTTGGDWQLETV-----Q 1369  
1196 QY LRQQFDTLREKREYNSPVKHYTDMGFTHN--KALEANYD---AVKAFINA----- 1240  
1370 Db TRDQIST--HDGRGSATSEHRLHGVSEVNVCGALTANVDNLTAVGANINATLEVQAQNI 1427  
1241 QY -----FKKEHHGVNLTTRTVLESQ---GSAELAKKLKNTL----- 1272



Db 1428 SLGAATDSLHTGESSKRHTSSVNLDETLIGSLQNLNATGIDINLQAAQDITLRASAVQTD 1487  
QY 1273 --LSLDSG-----EWSFSRSVGGGVSTVFV 1296  
Db 1488 GALTLAGGVDLLTQTQEHDEQRNHTGLSKGIASSTLTTRTDSLSOTLAVGSMLSAGSI 1547  
QY 1297 PTLSSKKVPFVPGAGITLDRAYNL-----SFSRTSGGLN 1331  
Db 1548 DVSGKNI---AVGSNVVAQDISLRAQENITVGTAAQSSSESHLFRQKXSLGMLSTGGIG 1604  
QY 1332 VSFERD-----GGVSGNIMVATGHVMPY-----MTGKKTIS 1362  
Db 1605 VTVGSSSTKMTDGGQSSISVSGSTVGSVLGNVMTAGEDLRVQGAELVLAGKDINLTGKNVS 1664  
QY 1363 AGNASDMSAKHKI--SPDLRIGAAVSTLQGTQLNSL--KFKLTDELPGFIHGLTHGTLT 1420  
Db 1665 ILAENGLTOSHVEVQKQSGILTALSAGVAVNVTATKAASESSGRLGAL----- 1718  
QY 1421 PAELLQKGIHQMKQKQSKLTFSDVTSANLDRAGINLNEDGS--KPNGVTVARVSAGLSAS 1478  
Db 1719 -----QGVKAALN-----GVQAVQAGQLVQAEAGGDAASWFGISAS 1753  
QY 1479 ANLAAGSRPSTTSGQ-----FGSTTSAGNNRPTFLNGVGAGAN----- 1517  
Db 1754 ---LGSQSSSEHQHQTHVTGSLTAGNNLTINATGEGNAANSBGDIVVQGSLOQAGGD 1809  
QY 1518 ---LTA--LGVAH-----SSTHEGKPGVFPAPTSTNVGAALALDNRKTSQISLEL 1564  
Db 1810 TTLDAARDVLLGNAANTQKTDGSSSSGGVGV-----SLGISGA-----SSGLSIFANA 1859  
QY 1565 KRAEPTVNDISLSTLTKHFKDSATTKMLAALKELDDAKPAQLHLOQHFSAKD--V 1622  
Db 1860 NKQSGSEHGDTGWTETT---LDSGTLTL-----YSGRDTSL 1894  
QY 1623 VGDE-----RYEAVRNKLXIVROAADSHSMELGSAHSSTYNNLS-----RINNDGIVE 1673  
Db 1895 VQAQVSGETVKVEGRDL--LLOSQDSNDYDAKQNSVGGFSFGSGMTGSIINGSD 1952  
QY 1674 LLHKHFDALPASSAKRGLGEMMNDPALKDIIQLQOSTPSSASVSNMELKDLREOTEKA 1733  
Db 1953 KLHNSFDSV-----QEQTGIFAGSGGFDITVGGHTQUDGA 1987  
QY 1734 ILGKVGRE---EVGVL--FQDRNN---LRVKSVSVSQSV-----SKSEGFNTPALLGT 1780  
Db 1988 VIGSTATADKNTLDTGTLGFSIDNQADFKVEHQSVGISTGGINIGSFVGNMANGLLVGA 2047  
QY 1781 SNS-----AAMSMERNIGTINFK--YQODONTPRFTLEGGIACANQVAVASALTDLX 1830  
Db 2048 NNEGHADSTTHAAVSE---GTITVRDNDQOQNVND---DLSRDRVEQAN--NALSPIFDKE 2099  
QY 1831 KEGLEMK 1837  
Db 2100 KEQNRLK 2106

## RESULT 12

Tl3564  
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)  
N:Alternate names: hypothetical protein EG:49E4.1  
C:Species: Drosophila melanogaster  
C>Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: Tl3564  
R:Spanos, L.; Papagannakis, G.; Siden-Kiamos, I.; Louis, C.  
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.  
A:Reference number: Z17689  
A:Accession: Tl3564  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-5327 <SPA>  
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1  
C:Genetics:

A:Cross-references: FlyBase:FBgn0025392  
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1  
A:Note: EG:49E4.1  
C:Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 2.5%; Score 240.5; DB 2; Length 5327;  
Best Local Similarity 18.4%; Pred. No. 0.008;  
Matches 370; Conservative 271; Mismatches 803; Indels 567; Gaps 81;

QY 18 AHPVHGVALQOQSSSSSPONAAASLAEGKNGKMPRIHQIPSTADGI-----SAAH 71  
Db 3517 ASRPTSAVESVQDEAEKSESRRESVAESKSLAK--EASRPASVAESVKDEAEKKEE 3574  
QY 72 QOKKFSRLGCLGTKKFSRPAQOQ---PGTTHSKGATLRLDLLARDGEGTQHEAAADAA 128  
Db 3575 SRRESVAESPLASKASRPASVAESVKDEAEKSESRRESVAESPLSKESARTSV 3634  
QY 129 -----RLTRSGVKRRNMDM--AGRPMTVKGSGEDKVPQTQKRHLQNNFGQM 174  
Db 3635 AESVKDEADKSESRRESGAESPLASMEASRPTSVAESVKDETEKSESRRESVTEK 3694  
QY 175 RQTMLSKMAHPASA--NAGDRLQHSPPHIPGSHHIEKEEPPVGSKATTAHADREVEAQE 232  
Db 3695 SPLPSKEASRPTSVAESVKDEAEK-----SKEESRRESVAESKSPASKESRPSVAE 3747  
QY 233 DDDSEFQOLHQORLARENERNPQPKL---GVATPISARFPQKLTAVAESVLEGTDTTQ 288  
Db 3748 SIKDEAEGTQE---SRRESMPESGKAESIKGQSSLASKETSRPDSVSVSKDETE--- 3801  
QY 289 SPLKPQ--SMKGGAGVTPLAUTLDKGKLOLQAPNPPALNTLLKQTLGKDTQHYLAHAS 347  
Db 3802 ---KPEGSAIDKSQVARSPEVAVS-----AKDEKSPHSRPSVADKSPD-----AS 3846  
QY 348 SDGSQHLILLDNKGHLFDIKSTATS-----YSVLHNSHP-----GEIKGLAQAGTGSVS 397  
Db 3847 KEASRSL-----SVAETASSPIEEGPRSTADLSPLNLGAEAKGLP---TLSSPI 3894  
QY 398 DGSGKISLGSQTSHNKYMLSQPGEAHRSLLTGWHQHPAGARPQESLRLHDDKTHIL 457  
Db 3895 DVAEGDFLEVKAESSRPVLSKPAEFSQPDGTGHTASTPVDASPVLEEVEVQ----- 3949  
QY 458 HPELGWQSGADKTHSQLSRQADGKLYALKDNR--TLQNLSDNKSSEKLVDKIKSVSDQ 515  
Db 3950 HTTSGV-----GATGATAETDLDTETKSETVTQKSETTLPETLTSKVES----- 3995  
QY 516 RGQVAILTDPGRHKMSIMPDLASPSHLSLHFA-----DAHQGLLHKSE--LEAQ 568  
Db 3996 -----KVEVLESSVKQVEEKVQTSVKQAEITVTDLSQLTKKSEQLTEIK 4041  
QY 569 SVAISHGRVLVADSEKGLFSAAI-----PKQGDGNELMKAMPQHALDEHFGHDHQSFF 624  
Db 4042 SVLDTN-----ISNVTNLFSTAVETIEKKVQDVTEKVIKATEH--VSEHVTITGESSTET 4095  
QY 625 HDDHGQNLALVKNNFRQQHACPLGN-----DHQFHGMNLTLD 661  
Db 4096 SOEKSSLDLGTFSLETHITTVGSPEFTVTCERDEPVLHDIKEEDEHFRFPPSDVDK 4155  
QY 662 ALVIDNQLGHLHNTNPEPHILDMGLSLALQSGK-----LHYFDQLTKGWTGAESDCKQ 716  
Db 4156 AAILPPQ--PMRPLSPREEEVAKIVADVAKVLKSDKDIITDIPDFDE-----RQ 4202  
QY 717 LKKGLDGAAYLLKDGVEKRLINQSTSSIKHGTVENVFSLPHVRNKPEPGDALQGLNKDDK 776  
Db 4203 LEEKLKSTADTEESDKSTRDEKSLSEISVKVEISEKSSPDQKSGISIEEKDKIQSEK 4262  
QY 777 AQMAVIGVKNYALATEKGDIFRSFQIKPGTQOQLERPAQTLSREGISGELKDIHVDHKQNL 836  
Db 4263 A-----QLRQGILASSRP-----ESVASQSPESVPSQS-- 4291  
QY 837 YALTHEGEVFPREAWQNGAESSSMHLALPQS--ESKLKSLDMSHEHKPIATFEDGSOH 895  
Db 4292 -AASHE-----HKEVEL-----SESHKAEKSSRPESVASQVKMDKTSRPSAST---SQF 4338





Qy	958	RVVSKPSE	R	I	V	A	F	N	P	T	S	T	P	P	I	K	N	A	A	V	A	T	O	H	Q	R	E	G	L	K	P	L	I	Y	E	M	O	G	A	L	I	K	O	D	A	H	1017										
Db	1530	R	K	A	R	I	K	B	E	G	H	I	-----	Y	L	E	T	D	K	D	-	K	N	V	D	-	L	T	A	S	E	L	K	G	N	T	-----	Q	I	K	A	1567															
Qy	1018	N	V	R	H	N	-	A	P	Q	P	D	L	S	K	L	E	T	L	D	L	G	E	H	G	A	E	L	L	N	D	M	K	R	F	-----	D	E	L	E	G	A	T	R	S	T	V	1065									
Db	1568	D	I	N	L	N	D	I	V	E	S	Y	K	Y	K	E	L	-	F	G	K	G	B	I	G	R	V	T	O	T	S	A	K	S	V	G	T	D	A	S	F	D	H	L	H	S	L	E	G	D	V	N	Q	1626			
Qy	1066	L	G	O	H	Q	V	L	K	N	G	E	I	N	S	F	K	P	S	P	G	K	A	L	-----	V	O	S	F	N	V	N	R	S	G	D	L	S	K	L	-	Q	O	A	V	1112											
Db	1627	T	G	S	N	L	K	A	N	R	T	T	G	V	V	K	G	D	F	N	T	K	A	G	D	L	P	H	R	O	I	D	T	V	T	S	G	T	V	T	S	A	S	G	G	Q	S	A	G	I	S	L	T	D	O	G	1686
Qy	1113	H	A	-	T	P	P	S	A	E	S	K	O	S	M	L	G	H	F	V	-----	S	A	G	V	D	M	S	H	Q	K	E	I	P	L	G	R	D	P	N	D	K	T	A	L	T	K	S	R	L	I	D	T	1167			
Db	1687	E	T	Y	N	K	T	A	T	A	G	A	N	A	D	V	T	N	F	M	K	R	T	E	T	S	T	H	R	N	S	E	F	-----	N	A	L	S	G	E	L	Y	V	M	G	K	1735										
Qy	1168	V	T	I	G	E	L	-----	H	E	L	A	D	K	A	K	I	V	-----	S	D	H	K	P	D	A	1190																														
Db	1736	A	D	I	G	G	V	D	I	N	R	D	V	E	I	K	T	P	E	E	I	A	A	B	Q	A	A	E	K	A	E	V	K	E	N	E	A	S	E	A	T	A	K	E	T	E	A	E	D	N	V	A	E	K	1795		
Qy	1191	D	O	I	K	-----	Q	L	R	O	F	O	T	L	E	K	R	E	S	N	P	V	K	H	Y	D	-	M	G	F	T	-----	H	N	K	A	-----	L	E	A	1229																
Db	1796	D	K	T	K	P	K	F	K	L	T	D	E	I	A	A	A	F	T	K	E	D	F	F	A	A	K	A	R	E	E	D	K	K	G	F	T	L	S	A	B	I	S	T	K	A	R	E	K	E	T	1855					
Qy	1230	N	Y	D	A	V	K	A	F	I	N	A	P	K	E	H	-----	H	G	V	N	L	T	R	T	V	L	E	S	O	G	S	-----	-	A	1262																					
Db	1856	T	Y	E	L	K	V	G	V	A	E	A	H	S	A	A	D	A	I	S	N	K	A	R	O	I	I	D	Q	N	G	L	K	Q	D	T	V	A	L	Q	A	S	D	V	L	N	A	T	G	D	L	A	1915				
Qy	1263	E	L	A	K	L	N	T	L	S	L	D	-	S	G	E	S	M	F	S	R	S	Y	O	G	-----	V	S	T	V	F	T	L	S	K	V	P	P	1306																		
Db	1916	G	A	S	A	K	L	F	E	L	T	T	E	K	S	R	G	S	D	G	R	S	I	L	G	R	I	N	L	A	A	R	G	D	I	T	L	N	N	V	E	T	T	E	N	S	H	L	S	L	K	A	R	D	N</		

```

RESULT 14
F90073      hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: F90073
M;Kuroda, M.; Ohra, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui,
ma, A.; Mizutani-Tai, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimura,
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758, MUID:21311952; PMID:11418146
A/Accession: F90073
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2271 <KB>
A/Cross-references: GDB:A000018; PID:g13702612; PIDN:BAB43752.1; GSPDB:GNO0149
A/Experimental source: strain N315
C/Genetics:
A/Gene: SA2447

Query Match          2.5%; Score 236; DB 2; Length 2271;
Best Local Similarity 18.2%; Pred. No. 0.0032;
Matches 295; Conservative 249; Mismatches 674; Indels 406; Gaps 61;

Qy    277 AESVLEGTDT-----TOSPLKPOSMLKGSGAGVTPLAVTLDKGKIQLADPNPALNTLL-- 330
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    576 APTVVTVGNQTIEVGKTMNPVLTTDNGTG-T-VNTVTGLPSG---LSYDS--ATNSIIG 629

Qy    331 -KOTLGKDIOHYLAHASDGSOHLLDNKHGLFDIKSTAT-SYSVLNHSHPCEIKGKLA 388
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    630 TPTKIQSVTVTV-----STDQANN-----KSTTTTFINVDVTAP-----TVT 668

Qy    389 QAGTGSVSVDGSKGISLGGTSQSHNKMTLSQPGEAHRSLLTGICWGPAGAARPOGESIR 448
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    669 PIGQSSEVVPISPIKI--ATODNS-----GNAVTVTVTGL-----PSSLTFD 710

Qy    449 LHKDKTHILHELGVWQSQADKOTHSQLSQAQDGK--LYALKONRITLQNLDNNKSSEKLVD 506
       ::||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    711 STNTNISGTPTNIGT-----STISIVSDASGNKTFTFFKEVTRNSMSDVSTSGSQ 764

Qy    507 KIKSYSDVRQGQVAIIITDPGRHKSMIMPDLDAESPESHLSLI-HPADAHQGLLHGKSEL 565
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    765 QSQSVSTSRADSSASTSTSG----SIIVSTSASTSKSTSVSLSDSVSAKSLSTSESNS 820

Qy    566 EAQSVAIISHRLVADVDEGKLFSAALPKQGDGNELMKAMPQHAFDHGHFHQISGFPH 625
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    821 VSSSTSTG---LVNSQSVSSMSGSVSK-----SDSVSTSTSN-----STEKSESLST 845

Qy    626 DRHQNLNALVKNFROQHACPLNDHFIPGWNIITDALVIDNLGLHTNPPEILDMDG 685
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    846 -----STLSDSITSNSN-----STLSDSITSNSN-----STEKSESLST 868

Qy    686 HLGLSALQEGLHYFPOLTKGMTGAESDCQKKGLDGAAYLLKDGVEKRLINIINSTSSI 745
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    869 TSDSLSTSTG---LSDSLSMYSTGSLSKXSQSLSTSIGS-----STSLASL 911

Qy    746 KHGTENVFSL-PHVRNKPPEPGDALOGLNKDDKAQAMAVIGNVKYLALTTEKDIRSFQIKP 804
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    912 SDSTSNIAISTSTSLSBASTSDSISINSIASQS-----ASTSKSDSQSTSL 961

Qy    805 GIQQ-----LRPPACTLSREG-I SGECLKDIHVHDHKONIYALTHGEVFHQPREAWONGA 857
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    962 STDSDKSMSTSELSNSTSTSGSVGSJS---IAASQSVSTSTSDS---MSTSEIVSDSI 1015

Qy    858 ESSSWHKIALPOSESKLKSLDMSHBHEKPIATPEDGSQHQLKAGGWH-----AYAAP 908
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    1016 STSG-----SLSASDSKSMVSSS-----MSTSQSGSTSELSDSQSTSDSDSKSLST 1066

Qy    909 ERGPLAVTSGGQTVFNRLMQGVKKVIPGSGLTVKLASAQTCGMTAEGRKVSSKFSERI 968
       |||:||||::||::||::||::||::||::||::||::||::||::||::||::||::||::
Db    1067 QSGSTSTSTSGASVRTSESQSTSGSMSAQSDMSISITFSFGDST--SDSKSASTASESI 1125

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:07:11 ; Search time 26.0314 seconds  
(without alignments)  
3676.525 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTDLKKEGLEWKS 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	244.5	2.6	3259	1 GOB1_HUMAN	Q14789 homo sapien
2	240.5	2.5	5560	1 SPEN_DROME	Q88x83 drosophila
3	232	2.5	1577	1 HLYA_PROMI	P16466 proteus mir
4	229.5	2.4	2541	1 TLN1_HUMAN	Q9V490 proteus mir
5	221	2.3	1608	1 HLYA_SERMA	P15320 serratia ma
6	219.5	2.3	2541	1 TLN1_MOUSE	P26039 mus musculu
7	218.5	2.3	5262	1 MLT2_HUMAN	Q14686 homo sapien
8	217.5	2.3	1957	1 SPOF_SCHPO	Q10411 schizosacch
9	214	2.3	1781	1 AK12_HUMAN	Q02952 homo sapien
10	205	2.2	2660	1 YEEJ_ECO57	Q8X8V7 escherichia
11	204	2.2	1447	1 GOA3_MOUSE	P55937 mus musculu
12	204	2.2	2845	1 APC_MOUSE	Q61315 mus musculu
13	202	2.1	5171	1 BPEA_HUMAN	Q94833 homo sapien
14	199.5	2.1	3590	1 FHAB_BORPE	P12255 bordetella
15	199	2.1	1191	1 CING_MOUSE	P59242 mus musculu
16	195.5	2.1	2319	1 AKA6_HUMAN	Q13023 homo sapien
17	198.5	2.1	835	1 INVO_PONPY	P14708 pongo pygma
18	197.5	2.1	1979	1 TRIA_HUMAN	Q15643 homo sapien
19	196.5	2.1	1902	1 P2P_LACPA	Q02470 lactobacill
20	196.5	2.1	5038	1 PCLO_MOUSE	Q9GYX7 mus musculu
21	195.5	2.1	4377	1 ANK3_HUMAN	Q12955 homo sapien
22	195	2.1	3321	1 PCN2_HUMAN	O95613 homo sapien
23	195	2.1	5938	1 MAC4_HUMAN	Q95pk2 homo sapien
24	194	2.1	1861	1 MAP2_RAT	P15146 rattus norv
25	194	2.1	3358	1 PGCV_MOUSE	Q62059 mus musculu
26	194	2.1	5430	1 MACF_HUMAN	Q9upn3 homo sapien
27	193.5	2.0	1902	1 P3P_LACLC	P15292 lactococcus
28	193	2.0	2238	1 GOA4_MOUSE	Q91vw5 mus musculu
29	191.5	2.0	2738	1 PGCV_RAT	Q9erb4 rattus norv
30	190.5	2.0	8545	1 ANCL_CABEL	Q9n4m4 caenorhabdi
31	190	2.0	2442	1 CEP2_HUMAN	Q9bv73 homo sapien
32	189	2.0	1902	1 P2P_LACLC	P15293 lactococcus
33	189	2.0	2175	1 HMCU_DROME	P10180 drosophila

RESULT 1  
GOB1\_HUMAN  
ID GOB1\_HUMAN STANDARD; PRT; 3259 AA.  
AC Q14789; Q14398;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)  
DE (Golgi complex-associated protein, 372-kDa) (GCP372).  
GN GOLGB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94187728; PubMed=7511208;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
RA Renz M.;  
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane  
RT protein (giantin).";  
RL Mol. Cell. Biol. 14:2564-2576(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94257116; PubMed=8198703;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
RA Renz M.;  
RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as  
RT target of antibodies in patients with rheumatic diseases and HIV  
RT infections.";  
RL J. Autoimmun. 7:67-91(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95100974; PubMed=7802676;  
RA Schda M., Miumi Y., Fujiwara T., Nishioka M., Ikehara Y.;  
RT "Molecular cloning and sequence analysis of a human 372-kDa protein  
RT localized in the Golgi complex.";  
RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges  
CC of the Golgi complex.  
CC -!- SUBUNIT: Homodimer; disulfide-linked.  
CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.  
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the  
CC autoimmune disease Sjogren's syndrome.  
CC -!- SIMILARITY: Belongs to the golgin family.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC of the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; X75304; CAA53052.1; -.  
CC EMBL; D25542; BAA05025.1; -.

#### ALIGNMENTS

DR PIR; A56539; A56539.  
DR PIR; 152300; 152300.  
DR Genew; HGNC:4429; GOLGB1.  
DR MIM; 602500; -.  
DR GO; GO:000139; C:Golgi membrane; TAS.  
DR GO; GO:000795; C:Golgi stack; TAS.  
DR GO; GO:0016021; C:integral to membrane; TAS.  
DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.  
KW Golgi stack; Antigen; Coiled coil; Transmembrane.  
FT DOMAIN 1 3235 POTENTIAL.  
FT TRANSMEM 3236 3256 POTENTIAL.  
FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).  
FT DOMAIN 48 593 COILED COIL (POTENTIAL).  
FT DOMAIN 677 1028 COILED COIL (POTENTIAL).  
FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).  
FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).  
FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).  
FT DOMAIN 2420 2423 POLY-GLU.  
FT DOMAIN 2993 2996 POLY-SER.  
FT CONFLICT 1 39 MISSING (IN REF. 3).  
FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).  
FT CONFLICT 1765 1765 D -> G (IN REF. 3).  
FT CONFLICT 2950 2950 H -> D (IN REF. 3).  
FT SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;  
  
Query Match 2.6%; Score 244.5; DB 1; Length 3259;  
Best Local Similarity 18.6%; Pred. No. 0.0024;  
Matches 389; Conservative 311; Mismatches 816; Indels 577; Gaps 92;  
  
QY 23 GHGVNLOQSSSPQNAASLAEGKRGKMPRIHQ--PSTAADGISAAHQKSFSLR 80  
DB 603 GEGIA-----PIKMKVLEDTGQDFPLMNEESSLPFAVEKEQASTHEQRTSEI- 652  
QY 81 GCLGKTFSSAPQGPQGTTHSKGATLRLDLDGGETQHEAAPDAAR----- 129  
DB 653 -----SLNDAGVELKS--TKQDGD-KSLSAVPDIGQCHQDELERLKS 691  
QY 130 -----LTRSGVKRRNDDWAGRMVKGSGEDKVPTQQRKHQNN-----FGQMR 175  
DB 692 QILELEINFHQAQIEYKNDLDEKA-----KEISNLNLQILBEFKKNDNNSAFTALSEER 746  
QY 176 QTMLSKM-----AHPASANAGDRLOHSPPHIPGSHHEIKESPVGSTSK 218  
DB 747 DQLLSQVKELSMVTELRAQVKQLEMLNLAERQRLDYES---QTAHDNLLTQIHSLSI 803  
QY 219 ATTAHADRVIEAQED-DDSEFQOLHQORLARENPFPQPKLGV-----ATPISARFQP 271  
DB 804 EAKSKDVKIEVLQNELDDVQLQFSEQSTLIRSLQSQLOKNESEVLEGAERVRISSKVEE 863  
QY 272 KLTAFAESVLEGTDTTQSPKLP-----QSMKKGAGVTPPLAVTLDKGLQLAPDN-- 322  
DB 864 LSAQLSQKELEITKMDQLLEKKRDVETLQOTIEKQQVTEISFSMTKRWQLNBEKFS 923  
QY 323 -----PPALNTLLKOTGLKDTQHYLAHASSDGSOHLLDNKGHLFDIKSTATSYV 374  
DB 924 LGVEIKTLKQNLNLSRAEAKKEQVEEDNEVSGLKQN-----YDEMSPAGQISK 974  
QY 375 LHNHPGEEI-----KGKLAQAGTGSVSDGSKGIS---LGSQTGHNTKMLSQP-- 421  
DB 975 EELQHEFDLLKKEQKRLQALINRKELLQVSRLEELANLKDSEKKEIPLSETER 1034  
QY 422 GEHRSLITGIWHPAGAARPOGESIRLHDDKIHLHPELVGVQSQADKUTHQSLSQADG 481  
DB 1035 GEVEEDKENKEYSEKCVTSKQCEIETYL---KQTISEKEVEL-QHIRKQLEELAAEEQF 1090  
QY 482 KLYALKDNRTLN-----LSDNKSSEKLVDKIKSYSDV--QRGOVAILTDTTPGR 528  
DB 1091 QALVKQNNQTLQKTNQIDLLQAEISENQA---IIQKLITSNTDASGDGVALVKET--- 1144  
QY 529 HRMSIMPLSDASPESHISLSLHFADAHQGLHKGKSELAQSAVSHGRLVVADSEGLKFS 588  
DB 1145 --VVISPPCTGSSE-----HWKPELEEKILALEKEK---EQLOKKLQE 1182  
  
QY 589 A-----AIPKQGD-----GNELKMKAMPQHALDEHFGHDHQISGFFHDDHGOALNVLVN 637  
DB 1183 ALTSRAILKKAQEKERHLREELKQKDDYRLQEQDEQSKENENIGDQLRLQIQVRE 1242  
QY 638 NF-----POOHACPLGNDHQHPGWNLTALVIDNQLGLHHTNPPEHLDMGLGSL 690  
DB 1243 SIDGKUPSTQDQSC-----SSTFG--LEEPFKATE--QHTQP-----VLE----- 1281  
QY 691 ALQEGKHLFPDQLTKGWTGAESDCKOLKGLDGAAY--LLKDGVEVKR-----LNINQSTSS 744  
DB 1282 -----SNLCPDWPSEDESALOGGTVAQIKQAQLKEIAEKVELELKVSTTSE 1331  
QY 745 IKHGTENVFSLPHVRNKPPEGDALQGLKDDKAQAMAVI GVNKYALTEKGDIRSFQIKP 804  
DB 1332 LTKKSEVFLQEQINK-----QGL-----EIES--LKT 1358  
QY 805 GTQOLRPPAOTLSREGISGELKDIHVHDHKNLYALTHEGEVFPQPRE---AWQNGAESS 861  
DB 1359 VSHEAEVHAESLQOKLESSQLTAGLEHLRELQPKDELKLSKKEEDVSYLSGQUSEK 1418  
QY 862 WHKLA-----LPQSESKLSLDSHEHKPIATFEDSGHQLKAGGHAYAAPERGPIAVG 916  
DB 1419 EAALTQIQTETIEQEDDLIKALHTQLEMW--AKEHDERIKQLQVELCEMKQKPEE---IG 1472  
QY 917 TSGSQTVFNLMOGVKGVIPGSLGVKLSAQGTGTMGAERKVVSKFSERI---RAYAF 973  
DB 1473 EE-----SRAKQIQIKL-----QAALISRKALKENKSLQEELSARGTIE 1514  
QY 974 NPTMTSTPRPKNAAYATOHGWQRE-----GLKPLYEMOGALIKQLDAHNVRHNAFPD 1027  
DB 1515 RLTKSL-----ADVESQVAQNKEDTVLGRALIQEERDKLITEMD-----RSLLENQS 1564  
QY 1028 LQSKLETLDLGEHGAELLNMDKFRDELEQSAIRSVTVLQHQGVKLSNGEINSEFKPSP 1087  
DB 1565 LSSCESLKLALSG--LTDKELVKIESLSKSKIAESTEWO--EKHKELQKEY---- 1615  
QY 1088 GKALVGSFVNVRGGDLKSLQQAIVATPPSAESKLQSMLGHFVSAGVDMSH-----OKG 1142  
DB 1616 -ELLOSQ-----ENVSEARIEQHVAVROEK-QELYGKLRSTEAANKETEKQJQEA 1667  
QY 1143 EIPLGRQRPNDKALTALSRILDTV-----TIGELHLEADKAK-----LVSDHKPDADQ 1192  
DB 1668 EQMEEMKEKMPAKSKQKQKILLEENDRLRAEVHPAGDTAKECMTLLSSNASMKEE 1727  
QY 1193 IKQLRQFDTLRK-----RYESNPVGHYTDGFTHNKALEANYDAVKAFINAF 1241  
DB 1728 LERVKMEYETLSKKFQSLMSEKDSLSEEVQDLKHQIE--DNVSKQANLEAT----- 1776  
QY 1242 KXHHGVNLTTRTVLSQSAELAKKLKNTLLSLDS-----GESMSFSRSYGGGVSTVP-- 1295  
DB 1777 EKHDNQTNTVEGTQSIPEGTE-----EQDSLMSMSTRPTCSVPSPSAKSNPAVSKDFSS 1831  
QY 1296 -----VPTLSKKVPVPVPIPGAGITLDRAYNLSFSRTSG-----LNVSGRGGV 1340  
DB 1832 HDEINNVLOQIDQKERI-----AGLEBEKQKNEFQSTLENEKNTLLSQISTKDGEL 1884  
QY 1341 -----SGNIMVATGHDVMPYMTGKTSAAGNASDWSAKHKISPDLRIGAASVGTLOG 1392  
DB 1885 KMLQEBVTQNNLNLQIQEELSRTVKLKTAEBEK-----DLEE 1924  
QY 1393 TLQNSLKFKLTEDELPGFHLGHTGLTTPALLQKGLHOMKQSKLTFSDV----- 1444  
DB 1925 RLMNQLA-----ELNGST--GNVCQDVTDQAIKNELLESEMKNLKRKCVSEBEKQOLVK 1377  
QY 1445 --TSANLDIR-----AGINLNEGSKPQNV-----TARVSAGLS 1476  
DB 1978 EKTQVSESEIRKVEIKIQGAQ--KEPQNKSHAKLOELLKEQKQEVKQLOKDCIRYQEKIS 2036  
QY 1477 ASANLAAGSRERTTSGQFGSTTSASNNRPTFLNGVAGANLTAALGVASHSTHEGKPVG 1536  
DB 2037 A-----LERTVKALEFVQTSQKDLKITTEN-----LAQAVEHRKKAQELAS 2079  
QY 1537 IPFAFTSTNVSAALAL-DN-----RTSQSISLSELKRAEPTVTSNDIS-ELASTLGHKF 1586



Chen F., Rebay I.;  
 "split ends, a new component of the Drosophila EGF receptor pathway,  
 regulates development of midline glial cells."; *Curr. Biol.* 10:943-946(2000).  
 [9]  
 RN FUNCTION ON WG PATHWAY.  
 RP MEDLINE=22668876; PubMed=12783785;  
 RX Lin H.V., Dorroquez D.B., Cho S., Chen F., Rebay I., Cadigan K.M.;  
 "split ends is a tissue/promoter specific regulator of wingless  
 signaling."; *Development* 130:3125-3135(2003).  
 RL  
 CC -1- FUNCTION: Probable corepressor protein, which regulates different  
 CC key pathways such as the EGF receptor and wg pathways. Involved in  
 CC neuronal cell fate, survival and axon guidance, cell cycle  
 CC regulation and repression of head identity in the embryonic trunk.  
 CC May act with the Hox gene Deformed and the EGF receptor signaling  
 CC pathway. Positive regulator of the wg pathway in larval tissues  
 CC but not in embryonic tissues. May act as a transcriptional  
 CC corepressor protein, which repress transcription via the  
 CC recruitment of large complexes containing histone deacetylase  
 CC proteins.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative promoter;  
 CC Comment=2 isoforms, 1 (shown here) and 2, are produced by use of  
 CC alternative promoters;  
 CC Event=Alternative splicing; Named isoforms=4;  
 CC Name=1;  
 CC IsoId=Q8SX83-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q8SX83-2; Sequence=VSP\_008565, VSP\_008566;  
 CC Note=No experimental confirmation available;  
 CC Name=3; Synonyms=Spn1;  
 CC IsoId=Q8SX83-3; Sequence=VSP\_008567;  
 CC Note=Produced by alternative splicing of isoform 1;  
 CC Name=4; Synonyms=Spn5;  
 CC IsoId=Q8SX83-4; Sequence=VSP\_008565, VSP\_008566, VSP\_008567;  
 CC Note=Produced by alternative splicing of isoform 2;  
 CC TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization  
 CC in stage 3 embryos, and in blastoderm cells, including pole cells.  
 CC Expressed throughout the rest of embryogenesis. Later, it is  
 CC expressed at higher level in epidermal cells and CNS.  
 CC -1- DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and  
 CC zygotically.  
 CC  
 CC -1- SIMILARITY: Belongs to the Spn family.  
 CC -1- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.  
 CC -1- SIMILARITY: Contains 1 SPOC domain.  
 CC -1- CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; AF182205; AAF13218.1; -  
 CC EMBL; AF184612; AAF26299.1; -  
 CC EMBL; AF221715; AAF34661.1; ALT\_INIT.  
 CC EMBL; AE003590; AAF51534.2; -  
 CC EMBL; AE003590; AAF51535.2; -  
 CC EMBL; AE003590; AAN10511.1; -  
 CC EMBL; AY094788; AAM11141.1; ALT\_SEQ.  
 CC HSP; P09651; IHA1.  
 CC FlyBase; FBgn016977; spen.  
 CC GO; GO:0007411; P:axon guidance; IMP.  
 CC GO; GO:0008347; P:glia cell migration; IMP.  
 CC InterPro; IPR000504; RNA\_rec\_mot.  
 CC Pfam; PF00076; rrm; 3.  
 CC SMART; SM00360; RRM; 3.  
 CC PROSITE; PS50102; RRM; 3  
 CC PROSITE; PS00030; RRM\_RNP\_1; FALSE\_NEG.

DR PROSITE; PS50917; SPOC; 1.  
 KW Transcription regulation; Repressor; Developmental protein;  
 KW Nuclear protein; Repeat; RNA-binding; Coiled coil;  
 KW Alternative promoter usage; Alternative splicing.  
 FT DOMAIN 554 632 RNA-BINDING (RRM) 1.  
 FT DOMAIN 656 730 RNA-BINDING (RRM) 2.  
 FT DOMAIN 734 806 RNA-BINDING (RRM) 3.  
 Query Match 2.5%; Score 240.5; DB 1; Length 5560;  
 Best Local Similarity 19.5%; Pred. No. 0.0085;  
 Matches 422; Conservative 256; Mismatches 744; Indels 739; Gaps 106;  
 QY 6 LGTEHKAHVTAHNPVGHGVALQQSSSSSSPQNAAS-----LAABG-----KNRGKMP 55  
 DB LVTVQANNSTLAN-----SNSSPSSVSASAVFATAAGGSSERSNRDRPY 129  
 QY 56 RIHQPTAAGDGSAAHQKKSF-----SLRGCLGTYK--FSRSAPQGGTTHSKGATL 107  
 DB RNSASVQGGGINSNTTTTAACTAGGSGGAIGTGTGLVGGPGGVP-----QAL 182  
 QY 108 RDLARDGDTGHEAA--AP-----DAARLTRGGVKKRRNMDMAGRPVKGSGEDKVP 160  
 DB GDRSSQNTQNHQHSARVAPQSWYEAATAATTAQLKS-----SGSG----- 225  
 QY 161 TQQRHQLNFGQMRQTMLSKMAHPASANAGRLQHP-----PHI--PGSHHFIKE 210  
 DB PVVGAGSCPSAAQGPQIQSQSTTAVHRSVAVAGSAADLLNTATSRNMLLHSSKLNKL 323  
 QY 249 -----ERENPPQPKLGVATPISARFQPKLTAVAESVLEGTDTTTOSPLKPO--- 294  
 DB LKAGATGSGGSGSGSGSPGACGAIPLTITTSNFSNSLNNTTITATPTMTIAS 383  
 QY 295 -----SMLKSGS--AGV-----TPLAVTLDKGLQLQAPNDPPALNTLLKQT 333  
 DB GAAGSVGLGSGAGAGVCSNSTAGSDILNVAALAAVDNG-----VPTHEIRTHNLH-- 437  
 QY 334 LGXDTQHYLAHASSDGSQHLNLDKNGHLFDIKSTATSYSLHNSHPGEIKGLAQAGTG 393  
 DB GRSTTSSSRSHSRSPSY-----SSSHSSSSSSSHSSHS--HASSPVQSGNCAMA--- 486  
 QY 394 SVSVDGKSGKISLGSGTQSHNKTMLSQPGEAHRSLLTGIWOHPAGARPQGESIRLHDDK 453  
 DB EGRSSR-TVNSVTVTSNS-----NPSGTA----- 510  
 QY 454 IHLHPGLVWQSGADKTHSQLSQADGK-----LYALKDNRTL-----ONLSDNKSSE 502  
 DB --VTVSAGVGGGCGSGSSSSSSSSSSSSSCLTANPVVHSEDNRPLAIRVRLPARSDDT 568  
 QY 503 KLVDKIKSYSDVGQOVA---ILTDTPGRHKM-----SIMPSLDASPESHISLSLHFA 552  
 DB SLKDGL-FHEYKHKGVTVVWVQNSERYALVCFKPPDDVEKALEVSHDKHFGGCKIEV 627  
 QY 553 DARQGL-----LHGKS-----ELEAQOVA---ISHGRVAVADSSGKL 586  
 DB EPIQGYDVEDNEFRPYEAELEDEYHPKSTRFLTIGNLEKIDITAGELRSH-----FEAFGEI 682  
 QY 587 FSAAIKQK-----DG-----NELKM---KAMPQHA--LDEHF 614  
 DB IEIDIKKQGLNAYAFQCYSDIVSVWAMRKMDGHEHLGNSRIKLGFGKSMPTNCVWID--- 739  
 QY 615 GHDDHQS-GFFHDDHQLNALVKNFRQHQHACPLGNDHQHPGWNLTDAIVINDQLGHH 673  
 DB GVDEKVSSEFLQSQFTFRFGAVTKVSDIR-----NRQLALVLVDQV----- 779  
 QY 674 TNPEPHEILDMHGLSLAQEGKHLHYFDQLTKWGTGAESDC-----KQLKGLDGAAYL 727  
 DB QNAQA-AVKDM--RGTI-LRRKKLV-----DFASRECQDAFYDKQEKQKQSS--- 824  
 QY 728 LKDGVEKRLINQSTSSIKHTGNTVSLPVRNKP-----PGDALQGLNKDKKA--- 777



```
Db 825 ---GSPFRFRYESSASSLQSRSSASSFRHQNNSNDCCSPINTPGASSGSISSASNLIN 881
Qy 778 QAMAV-----GVNKYLAITEKGDILRSQIKPGTQOOLERPACTLSREGISGELKDIHYDHK 833
Db 882 QSTISINIGTNCASAM-----PAPSLASAVVS-----CNVNAS 916
Qy 834 QNYALTHEGEVHFQPREAWONGAESWHKIALPQSESKLSLDMGHEHKPIATFEDGS 893
Db 917 GTVPAST-----SMPGVSSSS---SSLPMSPAL-----AQRHVMVNAQTQV 957
Qy 894 QHQLKAGWHAYAAPERGLAVGTSGQTVFNRL-----MQG-----VKGK 934
Db 958 DCDNEVGRFLFRSSEE---VSGGAGNSTOFEDVRCDSPVTARQGSVAVNCFGTPTAAVGE 1014
Qy 935 VIPSGSLTVKLSAQTGCTGAERKVKSPSERIRAVAFNPTMTSTPRPKNAAVATQHW 994
Db 1015 SIDGTLLNNQITGAEGFTSGGILSRRCCK-----TPK----- 1050
Qy 995 QGREGLKPLYEMOGALIKQDAHVNRHNAPOPDLQSKLETLDLGHEGAELNDMKRFRDE 1054
Db 1051 ----DLHPVHNRQLAEQVE-----ECFSSGDEG---VVSPRKRIKMD 1087
Qy 1055 LEQSATSVTVLGOHQGVKNSGNGINSEFKSPKALVQSPNV-----NRSGQLSKSLQ 1109
Db 1088 YHHHHHS-----NAGSVSTGEHSSINKDSP--LLLSNCDVIHDPILNR-----KSEI 1133
Qy 1110 QAVHATPPSAESKLSQMLGHFVSA--GVDMHQKGEIPLGRQRPDNDKLTAKRLILDT 1167
Db 1134 RRVSETSGSPS--IKFPGHLPSPAPQSLMLSCRRPSIDVGA-----LSLSSSSAP--- 1182
Qy 1168 VTIGELHELADKALVSDHKPDADQIKQLQROFDTLREKRYESNPVKHYTDMGTTHNKAL 1227
Db 1183 ----RHGIVGASSMDQHHMNASAAKRRVTTTMOQPSSSS-----ITNSSS 1226
Qy 1228 EANYDAVAFINAFKKEHGVNLTRTVLESGGAELAKLKNLLSLDSGESMSFSRSY 1287
Db 1227 GSGLLGSSSLTPA-DEVHHVSRGRGHQHLHSHSHEAS-----GGESADGSRP- 1273
Qy 1288 GGGYSTV-----FVPTLSKKYVVP-----VIPGAGITLDRAYNLSPSRTSG 1328
Db 1274 --GTPLCDEPEREVLFTPEPRLLPPPRVRERTRDVMMLPLPKFGVLF---FQQQSSRSSG 1328
Qy 1329 GLNVSGRGGVSGNIMVATGHDVMPYMTGKTSAGNASDWLSAKHKISPLDRIGAAVSG 1388
Db 1329 G-----GG-----AGNVIQQQLGGSGTGLGC-----IGAASSS 1358
Qy 1389 TLQGTQLNSLKFKLTEDLPFGIHLTHGTLTPAELLQKIEHQMKQSKLTFV-----D 1444
Db 1359 AC--SLNNS--SLNASQGMGSCSGSTPLSPSSRYWRSSHQVQNNHQOQQSQQLHGS 1413
Qy 1445 TSANLIDRA-----GINLNEGSKPENGVT-----RVSAG----- 1474
Db 1414 SSSNTCLMASPARPRSUNSSSDVPCQNAAGGSPSLDERLRNFENYERWGGSSREHI 1473
Qy 1475 -----LSASANLAAG--SRERSTTSQFGS-----TTSASNNRPTFLN----- 1510
Db 1474 SGHTPSSATPQWLSMHNLTGLNSHQTSSASGNSNSSSTGVSSASNSRHKFLDIDEL 1533
Qy 1511 -----GVGAGANL-TAALGVVAHSSTHEGK 1533
Db 1534 QPSDIVKSVLAKKSVFDDDFQRLNKNQWYDFSSSDFALGSSSNIVTGSSLVANVSRHPGG 1593
Qy 1534 PVGIFPATSTNVSAALADNRTSQSISLELKRAPVTSNIDISLTSGLKHKFDSATTK 1593
Db 1594 P-----CSGNTSPALP-----NLAATKATPIIGNSCGGLNGTGS--KSAGLLQ 1635
Qy 1594 MAAALKELDDAKPAQLHILOQHFSAKDVGVDERVEAYENLKKLVIRQQAADSHSMELGS 1653
Db 1636 RLSLSLPMN--SPQASMPFYNPSPSVGG-----VTACLGQITKPAAPGTASAGL-- 1685
Qy 1654 ASHSTYNNLRINNDGVELLHKHFDALP-ASSAKRLGEMMNDPALKDIIKOLQSTP 1712
```

---

```
Db 1686 -SGGTAASSSPAAANSQFTKGLQYPPFSPHPLENTAAPPVAPQAPPPLPMGKQSRUTG 1744
Qy 1713 FSSA---SVSMELKDG-----LREQTEK-AILDGKVGREEVGLFQDRNNLRVK-SVSVS 1762
Db 1745 QSSGNLTKSLVDPDGPQSSPARVOLQKSASVPGSTNVGAPSSLSLDTTASVTSASIS 1804
Qy 1763 QSVSKSEGFNTPAILL-----LGTNS- AAMSMERNIGTINFKYGQDQ 1803
Db 1805 SSTSGNSSLTSAAIHVQKPPQSTFVEEHTKKSGTSTQSSSSSSSKKISSTHDKLHSHK 1864
Qy 1804 N 1804
Db 1865 N 1865

RESULT 3
HLVA PROMI
ID HLVA PROMI STANDARD; PRT; 1577 AA.
AC P16466;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE Hemolysin precursor.
GN HPMA.
OS Proteus mirabilis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Proteus.
OX NCBI_TaxID=584;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-43.
RC STRAIN=Isolate 477-12;
RX MEDLINE=90170827; PubMed=2407716;
RA Uphoff T.S., Welch R.A.;
RT "Nucleotide sequencing of the Proteus mirabilis calcium-independent
RT hemolysin genes (hpmA and hpmB) reveals sequence similarity with the
RT Serratia marcescens hemolysin genes (shlA and shlB).";
RL J. Bacteriol. 172:1206-1216(1990).
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood
CC cell membranes and cause cell rupture by mechanisms not clearly
CC defined.
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
CC ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. HPMA
CC REQUIRES HPMB FORMATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- MISCELLANEOUS: THE CONSERVED AMPHIPATHIC DOMAINS IN SHLA AND HPMA
CC MAY BE RESPONSIBLE FOR PORE FORMATION.
CC -!- SIMILARITY: TO S.MARCESCENS HEMOLYSIN (SHLA).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC PIR: M30186; AAA25657.1; -.
CC DR A35140; A35140.
CC DR InterPro; IPR008638; Haemagg_act.
CC DR Pfam; PF05860; Haemagg_act; 1.
CC KW Hemolysins; Toxin; Outer membrane; Signal.
CC FT SIGNAL 1 29
CC CHAIN 30 1577 HEMOLYSIN.
CC SEQUENCE 1577 AA; 165869 MW; 175975E0C924B2D9 CRC64;
CC
CC Query Match 2.5%; Score 232; DB 1; Length 1577;
CC Best Local Similarity 18.9%; Pred. No. 0.003;
CC Matches 329; Conservative 238; Mismatches 614; Indels 564; Gaps 84;
CC
CC Qy 261 VATISARFQKLTAESVLEGTDTTQ-----SPLKPSQMLKSGAGVTFPLAVTLQK 314
CC Db 197 IAPRIDGR--GKITAABISAFTQNTFSQHPDIUSSQKPVSAIDSYFFG-----SMQSG 248
```



CC similarity).  
 CC -|- SUBUNIT: Binds with high affinity to vinculin and with low  
 CC affinity to integrins (By similarity).  
 CC -|- SIMILARITY: Contains 1 PBM domain.  
 CC -|- SIMILARITY: Contains 1 I/LWEQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL; AF078828; AAD13152.1; -  
 CC EMBL; AF177198; AAF23322.1; -  
 CC EMBL; AF178534; AAF27330.1; -  
 CC EMBL; AF178081; AAF27330.1; JOINED.  
 CC Genew; HGNC:11845; TLN1.  
 CC MIM; 186745; -  
 CC GO; GO:0005925; C:focal adhesion; NAS.  
 CC GO; GO:0005200; P:structural constituent of cytoskeleton; NAS.  
 CC GO; GO:0006928; P:cell motility; NAS.  
 CC GO; GO:0007016; P:cytoskeletal anchoring; NAS.  
 CC InterPro; IPR00299; Band 4.1.  
 CC InterPro; IPR002558; ILWEQ.  
 CC Pfam; PF00373; Band 41; 1.  
 CC Pfam; PF01608; ILWEQ; 1.  
 CC ProDom; PD011820; ILWEQ; 1.  
 CC SMART; SM00295; B41; 1.  
 CC SMART; SM00307; ILWEQ; 1.  
 CC PROSITE; PS00660; FERM 1; 1.  
 CC PROSITE; PS00661; FERM 2; 1.  
 CC PROSITE; PS00657; FERM 3; 1.  
 CC PROSITE; PS00945; ILWEQ; 1.  
 CC Structural protein; Cytoskeleton.  
 CC DOMAIN 86 403 FERM.  
 CC FT DOMAIN 2340 2533 I/LWEQ.  
 CC FT CONFLICT 824 824 G -> R (IN REF. 2; AAF27330).  
 CC FT CONFLICT 1227 1227 S -> L (IN REF. 1).  
 CC FT CONFLICT 1549 1549 P -> A (IN REF. 2; AAF27330).  
 CC FT CONFLICT 1604 1604 Q -> K (IN REF. 2; AAF27330).  
 CC FT CONFLICT 1701 1701 E -> Q (IN REF. 2; AAF27330).  
 CC FT CONFLICT 1718 1718 H -> N (IN REF. 2; AAF27330).  
 CC FT CONFLICT 1966 1966 A -> R (IN REF. 1).  
 CC FT CONFLICT 2256 2256 MISSING (IN REF. 2; AAF27330).  
 CC SEQUENCE 2541 AA; 269717 MW; 47PD9B22BCF47296 CRC64;  
 Query Match 2.4%; Score 229.5; DB 1; Length 2541;  
 Best Local Similarity 19.5%; Pred. No. 0.0083;  
 Matches 449; Conservative 297; Mismatches 846; Indels 707; Gaps 114;  
 QY 11 KAAVHTAAHPVG---HG-VALQ--QSSSSSPON-----AAASLAEGKRGKMPR 56  
 DB 428 KSTVLOQYNRVGVHGSVALPMSRSGASGPFVQSGMPPAQQITSGMERGHMP 487  
 QY 57 IHQPSTADG-----ISAHQOKSFLRGLG---TKFSRAPOGPGTTHSK- 103  
 DB 488 LTSQAQALGTINSSMQAVQAQATLDDFDLPGLQDAASAKANRKNKDESKHHSQV 547  
 QY 104 -----GATURDLARDGETOHE-----AAPDAARLTR--SGGVK--RRNDDM--AGRP 148  
 DB 548 DAITAGTASVNLTAGDPAEDTYAVGCAVTTISNLTMSRGLKLLAALLEDEGGSGRP 607  
 QY 149 MVKSGGEDKVPYTOQRHQLNFGMRQTMLSKVAHPASANAGRLQH-----SPPHIPG 203  
 DB 608 LLQAAGKLAGAVSELLRSQAAPASAPREPQNLQAAGNVGQA--SGELLQIGESDTPHPQD 666  
 QY 204 SHHEIKPEPVGTSKATTAHADRVETIAQEDDDSEFQHLHQRLARERNPQPKLGWAT 263  
 DB 667 ALMOI--AKAVASAAALVLKAK--SVAQRTSDSLQ-----TQVIAATQCALSTSLQVACT 720  
 QY 264 PISARFOPKLT-----VAESVLEGTDTTQSPKPSMLKSGAGVTPLAVT 310

DB 721 KWVA---PTISSPVQEQQLVEAGRLVAKAVEGCVSQAATEDGQLLRGVGAAT--AVT 775  
 QY 311 LDRKGLQLADPNPALNTLLKQTLGKDTQHVLAHASSDGSQHLDDNKGHLFDIKSTAT 370  
 DB 776 -----QALNELL-----QHVKAHATGAGPAGR-----YD-QATDT 804  
 QY 371 SYSVLHN-----SHPGEIKG---LAQAGT---GSVSDGKSGKISLGSGTQSHKNTMLS 419  
 DB 805 ILTVTENIFSMGMDGAGVMGQARILAQATSDLVNAIKADAE-----GESDLENSRKL 858  
 QY 420 QP---GEARSLTLGTIWOHPAGARPOGESIRLHDDKIHILHBPGLGWQADKDTHSOLS 476  
 DB 859 AAKILADATATKVV---EAAKGA-----AHPD-----SEBQQQLRL 891  
 QY 477 ROADGKLYALKDNRTLONLSDNKSSEKLVDKIKSYVDQRQGVAILTDTTPGRHKMSTPS 536  
 DB 892 EAAEG---LRMATNAAQ-----NAKKKLVRLE-HAAKQAAASATQITIAAQAAS-TPK 943  
 QY 537 LDASPESHISLSLH-FADAHQGLLHGKSELEAQVAISHGRLVVADSE-----GKLFS 589  
 DB 944 ASAGPQLLVQSCKAVAEQIPLLVQGVRSQAQPDSPSAQLALIAASQSFQPGKMWAA 1003  
 QY 590 A---IPKQGCNENLKMAMPQHLDHFGHDHQISGFFHDDHGQNLALVKV----- 637  
 DB 1004 AKASVPTIQD-----QASAM-----QLSOCAKNLGTALAE 1035  
 QY 638 --NFRQOHAC-PLGNDHQHFGWNL-----TDALVIDNQLGLHHTNPPEHIL--D 683  
 DB 1036 TAAQAQACGPLEMDSALSVVQNLKDLQEVKAAARDGKU-----KPLPGETWEKTQD 1090  
 QY 684 MGH-----LGSALQEGKHLHYFDQLTKGWTGAESDCQKQKGLDGAAYLLKDG 731  
 DB 1091 LGNSTKAVSSAIAQLLGEVA--QGNENYAGIAARDVAGG---LRSLAQAARGVAALTSDP 1145  
 QY 732 EVKELNINQSTSSIKHGTENVFSPLHVRNKPEDGALQGLNKDKQAQAMAVIGNKYL-- 789  
 DB 1146 AVQAIVLDTASDVLDKASSLIEEAKKAAGHPDPEQQRLAQAVAKVTQA---LNRVCSC 1202  
 QY 790 -----ALTEKGD---LRSPQIKPGTQQLERPACTLSREGISG-----ELKDIHV 830  
 DB 1203 LPQORDVDNALRAVDGASKLLSDSLPSTGTPOE-AQSRLENAAGLNGAATLQVQASR 1261  
 QY 831 DHKQNL-YALTHEGEVFPQPREAMQNGAESSSWHKLALPOSE-----SKLKSIDMSHEH 883  
 DB 1262 GTPQDLARASGRFGQDFSTFLEA--GVEMAG---QAPSQEDRAQVVVNLKLGISMSSK 1314  
 QY 884 -----KPIATFEDGSQHLKAGWHYAAAPERGLAVGTSGSQTV----- 923  
 DB 1315 LLLAAKALST--DPAAPNLKS---QLAAAAA---AVTDSINQLITMCTQQAQPGKECDN 1365  
 QY 924 -----FNRLMQGVKGVIPGSGLT-VKLSAQTG---GMT 953  
 DB 1366 ALRELETVRELLENVPQINDMSYFGCLSDVMSNKGVLGAMTGISQAKNGNLPFGDA 1425  
 QY 954 GAERKYSKFSERIRAYAFNPTMTSRPIKNAAYATQHGWQREGKPLPYEMOGALIKQ 1013  
 DB 1426 ISTASKALCGFTEAAQAAYLVGSDP-----NSAQOQGLVEPTQPARA--- 1470  
 QY 1014 LDANNVRHNPAPQDLQSKLETLDLGEHG---AELLNDMKRPRDELSQATRSVTVLGCHQ 1070  
 DB 1471 -----NQAIQMACQS-----LGEFGCTQAQVLS-----AATIVAKHT 1502  
 QY 1071 GVLKSNGEINS--EFKPSPGKALVQSP-NVNRSQDLSKSLQ-----QAVHAT 1115  
 DB 1503 SALCNRLASARTTNTAKRQFVQSAKEVANSTANLVTKITKALDGPFTTEENRAQCRAT 1562  
 QY 1116 PPSAE-----SKLQSMGLGH-----FVSAGVDMSHQKGIPLGR----- 1148  
 DB 1563 APLLEAVDNLSAFASNPFFSSIPAQISPEGEAAMEPIVISAQTMLESAGGLIQTARALAV 1622  
 QY 1149 -QRDPNDKALT-KSRILIDTV-----TIGELH-ELA-----DKAKLVS 1184

Db 1623 NRPDPSPSVLGHSTRTSDSIKKLITSMRDKAPQLECEETAALNSCLRDLDQASLAA 1682  
QY 1185 -----DH-----KPDADQ-----IKQLRQED--T 1202  
Db 1683 VSQOLAPREGISQSEALHTEMLTAVQESHILIEPLAHAARAEASQLGHKVSQMAQYFEPLT 1742  
QY 1203 LREKRYESNEPVKHYTDMGFTH-----NKALEANYDAVKAFINAFKKEHGVNLTTRTVL 1256  
Db 1743 LAAGAASKTSLSPQOMALLDQTKTALAESALQLLYTAKEAGGNPKQAAH-----TQEAL 1796  
QY 1257 ES--QSSAELAKKLKNTLSLDSEMSFSRSGGVSTVFTVTLAKKVPVPIPGAGIT 1314  
Db 1797 EEAQVMTEAVEDLTTTL-----NEAASAGVVGWMDSI--TQAINQLDRGPMGEPSGF 1850  
QY 1315 LDRAYNLSPRTSGGLNVSGFRDGGVSGNIMVATGHDVMPYMTG----- 1358  
Db 1851 VD--YQTTWRTAKAIVTQVE-----MTKNTSPSEELPLANQUTSYGLASE 1899  
QY 1359 ---KKTSGNASDNLGSAKHKISPLDRIGAAVSGTLQGTQNSLKFKLTDELPGFIHGLT 1415  
Db 1900 AKPAVAEAENEETIGSHIKHRVQ--ELGHGCAALVTKAGALQCSFSDAYTKKEL----- 1950  
QY 1416 HGLTTPAELLQKIEH--OMKQSKLTFSVDTSAN-----JDLRAGI-----NINED 1460  
Db 1951 ---IECARRVSEKVSHVLAALQAGNRGTQACITAAASAVSGIIADLDTTTFATAGTLNRE 2007  
QY 1461 GSKP-----NGV--TARVSA--GLSASANLAAGSRSTTSQO-----FGST 1498  
Db 2008 GTTFADHREGILTKAKVLVEDTKVLVONAGSQEKLAAQAQSVATITFLADVVLGAA 2067  
QY 1499 TSANNRPTFLNGVAGANITAAALGVAHSSTH--EGKPVGIFPAFTSTNVNSAALADNRT 1556  
Db 2068 SLGAEDPETQVVLINAVKDVAKALGDLISATKAAAGK--VGDDPAVQLKNSAKVMVTNVT 2126  
QY 1557 S-----QSLSEL-----KPAEPTVNDISELTSTLTKHP 1586  
Db 2127 SLKTKVAVEDEATKGTTRALEATTEHROELAVFCSPPEPAKTSTPDEFIRMTKGI----- 2182  
QY 1587 KDSATTKMLA-----ALKELDDAKPAQOLHLOQHFSKADVVGDERVEAVRNLK 1635  
Db 2183 -TWATAKAVAGNSCRQEDVIATNLSRRAIADMLPCKAEAAVHPVAPVRLRALHYGR 2241  
QY 1636 KLV-----IROQAADSHMELGSASH-----S 1657  
Db 2242 ECANGYLELLDHVLLTLQKPSPELQKQTLGHSKRVAGSVTELIQAAEAAMKGTBWDPEDP 2301  
QY 1658 TTYNNLSRINNDGIVELLHKKHFDAAALPASSAKRLGEMMNNDPALKDIIKQLQSTPSSAS 1717  
Db 2302 TVIAENELIGAAAEAAKKBOLKPRAPKPEADESLNFEQIILEAKSI-----AAA 2355  
QY 1718 VSMELKDGLREQTEKATLDGKVGREEVGVLFQDRNNLRVKSVSQSVSKS----- 1768  
Db 2356 TSALVKAASAAQRE--LVAQKVGAI PANAL--DDGQWSQGLISAARVVAATNNLCEAN 2412  
QY 1769 ---EGENTPALLGTSNAAAMSMERNIGTINFKYGGQDQNTPRPTLEG--GIAQAN----- 1819  
Db 2413 AAVQGHASQBLISSAKQVAASTAQLIIVACKVKADQDSEAMKRLQAAGNAVKASDNLVK 2472  
QY 1820 -POVASALITDKKEGLEMK 1837  
Db 2473 AAQKAAAFEEQENETVVVK 2491

RESULT 5

HLVA\_SERMA

ID HLVA\_SERMA STANDARD; PRT; 1608 AA.

AC PL5320;

DT 01-APR-1990 (Rel. 14, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-NOV-1990 (Rel. 16, Last annotation update)

DE Hemolysin precursor.

GN SHA.

OS Serratia marcescens.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Serratia.  
OC NCBI\_TaxID=615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.  
RC STRAIN=SN8;  
RX MEDLINE=88257037; PubMed=3290200;  
RA Poole K., Schiebel E., Braun V.;  
RT "Molecular characterization of the hemolysin determinant of Serratia  
marcescens";  
RT J. Bacteriol. 170:3177-3188(1988).  
CC -!- FUNCTION: Bacterial hemolysins are exotoxins that attack blood  
cell membranes and cause cell rupture by mechanisms not clearly  
defined.  
CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM  
ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHA  
REQUIRES SHLB FUNCTION.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; M22618; AAA50323.1; --  
DR PIR; A28182;  
DR InterPro; IPR008638; Haemagg\_act.  
DR Pfam; PF05860; Haemagg\_act; 1.  
KW Hemolysis; Toxin; Outer membrane; Signal.  
FT SIGNAL 1 30  
FT CHAIN 31 1608 HEMOLYSIN  
SQ SEQUENCE 1608 AA; 165078 MW; D869B476FE7DAD51 CRC64;  
  
Query Match 2.3%; Score 221; DB 1; Length 1608;  
Best Local Similarity 18.6%; Pred. No. 0.01;  
Matches 372; Conservative 235; Mismatches 689; Indels 700; Gaps 89;  
  
QY 33 SSSSPONAAASLAAGKRNKMPRIHPSTAA-----DGISAHQKKSFLR--GC 82  
Db 23 AASAGAYAAIAVANGANG---PGVSTAATGAQVDIVAPNGNGLSHNQYQDFNVNPGA 79  
QY 83 L--GTKKFSRSPQGPQGTTHSKG---ATLRLDILLRDDG-----ETQHEAAA 124  
Db 80 VLNSREAGLSQLAGQLGANPLNGREASVILNEVIGRNPSLLHQOEIPGMADYVLAN 139  
QY 125 PDAARLTRSGGVKRRNMDMAGRPVYKGGSGEDKVPYQKKRHQNLNFGQMRQTMLSKMAH 184  
Db 140 PNGISQSCGFIINTSHSLVVGNNFLVGVNGVLOG-----YSTFGNRTLSLN-- 185  
QY 185 PASANAGDLRLQSPPHIPGSHHEIKEEPPVGSSTKATTAHADRVET--IAQEDDDSEFQOLHQ 243  
Db 186 -GTINAGGVLDLTAPKIDSRGEVIVQDFKQSNQKVTSAAINALISGLNRVARDGTVQASQ 244  
QY 244 QRLAREE-----NPPQPKLGVAIPISARFOPKLTAV-----AESV 280  
Db 245 MPTALDSYILGSMQAGRINIINTAQSGVKGASLGNAGDELKVKADIRSERVDASSN 304  
QY 281 LEGTDTTQ-----SPLKPSMLKG-----SGAGVTPPLAVTLDKG 314  
Db 305 KNGGDNYQYRGGIYVNDRSSSTLTTRTELKGNISLVADNHAHLTATDIRGEDITLQGG 364  
QY 315 KLOLAPDNPALNTLLKQTLKDTQHYLAHSSDGSCHLLDNKGLHLPDIKSTATSYSV 374  
Db 365 KLTLL-----DQOQ--LKOTQGHDTDRWFSWOYDV 392  
QY 375 LHNHPGEIKGKLAQAQAGTSVSDGSKGIS-----LGSGTQSHNKTMLSQPSGAH-- 425  
Db 393 TRE-----REQLOQAGS--TVAASGSAKLSTOEDVKLLGANVSADRALSVAARDVHLA 445

QY 426 -----RSLTGIWQH-----PAGAAPOGE-----SIRLHDDK 453  
 Db 446 GLVEKDKSSRGYQRNHTSSIRLGRWNSDESESLKASERSEGETLTKAGRNVTQCAK 505  
 QY 454 IHLHPGLGWQSDAKDTHSOLSHQADGKLYALKDNRTLO---NLSNKSSEKLVDKIS 510  
 Db 506 VHAQRD-----LTIDAQNIQVGVQKATANAKAVRDRDKTSMGGIGGDNKNS-----552  
 QY 511 YSVDRQGVAILTDPGRHKNMPSLDASPEHSISLHPADAHQGLLHGKSELEA-- 568  
 Db 553 -----NREISHASELTSGLTLRL-----NGQGVTTITGSKARGQK 589  
 QY 569 -SVAISHGRVLVADSEKLSAALPKQDGNELKAMKAMPCHALDEHFGHDHISGFHDD 627  
 Db 590 GEVATHGGLRIDNA-----LSTTVDK-----IDARTGTAFNITSSSHK- 628  
 QY 628 HGQLNALVKNFRQOAHACPLGNDOHGFHGNLT-----DALVIDNOLGLHHTNPEPHEIL 682  
 Db 629 -----ADNSYQSSTASELSKD-----TNLTILVSHKDAVDIGSOVA-----663  
 QY 683 DMHGLGSLALQEGKHLHYFDQITKGWTAESDCKQLKGLDGAAYLLKDGVEKVRNLINQST 742  
 Db 664 -----SGGE-----LSVESKT 674  
 QY 743 SSIKHGTENVFSLPHVRNKPPEPGDALQGLNKDDKAQAMAVTGVNKLALTEKGD-----796  
 Db 675 GNI-----NVKA-----AERQONIDEQKALTNGYAK-----EAGDKQYRAG 712  
 QY 797 --IRSFQIKPGTQOLERPACTLSREGISGELKDIHVDHKQNLIALYTHEGEVFGHPREAWQ 854  
 Db 713 LRIEHTRDSSEKTRTENSASLS--GGSVLK-----AEKDVTFSGSKLVAD 757  
 QY 855 NGAESSWHKIALPQSSSKLSKSDMSHEKPIATFEDGSQHLKAGGWHAYAAPERGFLA 914  
 Db 758 KDAVASGNKVSLAADKTAS-----NTEQTKIGGFY-----792  
 QY 915 VGTSGSTVFNRLMQGVKGVIPGSGLTVKLSAQGTGWTGABGRKVSFKSERIRA--Y 971  
 Db 793 --TGG-----IDKLGSGVE-----AGYENKTKQAQSSKAITSGSDV 826  
 QY 972 AFNPTMSTPPIKNAATATQHWQREGKPLIYEMQALIKQLDAHNVRHNAPODLOSK 1031  
 Db 827 KGNLTINARDKL-----TOGGAQ-----HSGVGAY--QENAAQVGDHIAAADTASTT 870  
 QY 1032 LETLDLGEH--GAELNDMKRPRDELEOSATRSVTLVGHQGVKLSNGEINSEFPKSPGKA 1090  
 Db 871 TTKTDVGVNIGANV--DYSATVRPVERAVGAAKL--DATGVINDIGIGA--PNVG-- 921  
 QY 1091 LVQSFNVRNSQDLSK--SLQQAQVHATPPSAESKLSQMLGHFVSAGVDMSHQKEIPLGR 1148  
 Db 922 ---LDIGAQGSSEKRSSSQAVVSS-----VQAGSIDINAKGEV--- 958  
 QY 1149 QRPNDKXTALTKSLILDVTIGELHELADKALVSDHKPDADQIKQLRQFPDITREKRY 1208  
 Db 959 -RDQGTQYQASK-----GAVNLTAQ-----SHRSEAAANRQDEQSRDT-----R 996  
 QY 1209 ESNPVKHYTDMG-----FTHNKALEANYDAVKAFINAFKPK-----HHGVN 1249  
 Db 997 GSAGRVYVTTGSDLTVDKAGEGQTRSNSSASQAVTGSIDAANGINUNVKDAIYQGT 1056  
 QY 1250 LT---TRTVLESQSAEL-----AKKDKNTLLSLDSG-----ESMSFSSYSGGV--- 1291  
 Db 1057 LNGRGKTAVNAGGDIRLDAQSDKQSESRSGFNFKASAKAGFTADSKNFGAGFGGTHNG 1116  
 QY 1292 ---STVFVPTLSKVPVPIPGAGITLDRAYNISFSRTSGGLNVSG-----1335  
 Db 1117 ESSSTAGVGNISGQGVVELKAGRDLTLQ-----TDVKSQGDVSLSAGNKVALQAESTQ 1172  
 QY 1336 --RDGVSGNTWATGHVDMVPMYTKTKTSAGNASD-----WLS 1371  
 Db 1173 TRKESKUSGNDILGAGSSDSEKKTGNLSAGAFDIAKNVESATERQATIASDKVTLIS 1232  
 QY 1372 AKHKISPDLRI--GAAVSG---TLQ-----GTLQNSLKFKLTEDELPGFIHGLTHGLTPAE 1423

Db 1233 ANGKGDLDALHLOKAVKSGGALEAKNGGILLSEAKNEQHKDN-----1275  
 QY 1424 LLQKGIHQMKQSKLTFSVDTSANLD-----LRAGINLEDSGSKN-GVTA 1469  
 Db 1276 -WSLGIKANAKGGQ--TFNKDAGKVDPTNGKDTHTLGLAGLVGVQEQDKTTHANTGITA 1332  
 QY 1470 -RVSAGLSASANLAAGSRERSTTSGQFGSTTSASNRRPTFLNGV-----GAGA 1516  
 Db 1333 GDVTLSNGKDTRLAGARVDADSVQGVGGDLHV--ESRKDVENGVKVDVDAGLSHSNDPGS 1391  
 QY 1517 NLTAALGVAIHSSHEGK-----PVG-----IPPAFTSTNVSAALADNRTSQSISLELKRAE 1568  
 Db 1392 SITSKLSKVGTTPRYAGKVEKLEAGVNVKADATTDKYNVSVARRLDPOQDITGAVSFSAE 1451  
 QY 1569 -----PVTSNDISEL-----TSTLGHKFKDSATTMKAALKELDKAPAEQLHILQOH 1616  
 Db 1452 GKVTLPATPAGERKPGFLWDRGARTVGGAVKDSIT-----GPAGR-----QGH 1494  
 QY 1617 FSA-----KDVVGDERYEAVRNLKKLIVRQQA-----ADSHSMELGSASHSTTVNN 1662  
 Db 1495 LKNVADVNNNAVGEQSAIAGKNGVALQVGQOTLTGTGEIRSQQKVELGGSQVSDQVN 1554  
 QY 1663 LSRINDGIVELLKHFDALPAPASSAKRGLCEMNDPALKDIIKLOLSTPSSASVSMEL 1722  
 Db 1555 GORYQGGGRYD-----AAATVGGLLGG--AAKQSV--AGNVPPFASGHASTQQ 1597  
 QY 1723 KDGLREQTEKAILDGK 1738  
 Db 1598 AD-----AKAGVFSGK 1608

RESULT 6  
 TLN1 MOUSE  
 ID TLN1 MOUSE STANDARD; PRT: 2541 AA.  
 AC P26039; Q8VEF0;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DB Talin 1.  
 GN TLN1 OR TLN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=Fibroblast;  
 RX MEDLINE=91015390; PubMed=2120593;  
 RA Rees D.J.G., Ades S.A., Singer S.J., Hynes R.O.;  
 RT "Sequence and domain structure of talin.";  
 RL Nature 347:685-689(1990).  
 RN [2]  
 RP SEQUENCE OF 1603-2541 FROM N.A.  
 RC TISSUE=Mammary gland;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.I., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.B.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length  
 RL human and mouse cDNA sequences.";  
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Probably involved in connections of major cytoskeletal  
 CC structures to the plasma membrane. High molecular weight  
 CC cytoskeletal protein concentrated at regions of cell-substratum  
 CC contact and, in lymphocytes, at cell-cell contacts.  
 CC -!- SUBUNIT: Binds with high affinity to vinculin and with low  
 CC affinity to integrins.  
 CC -!- PTM: Phosphorylated.  
 CC -!- SIMILARITY: Contains 1 FERM domain.  
 CC -!- SIMILARITY: Contains 1 I/LWEQ domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X56123; CAA39588.1; -;  
 CC EMBL; BC018557; AAH18557.1; -;  
 CC FIRM; S11661; S11661.  
 CC MGD; MGI:1099832; Tln.  
 CC GO; GO:0005925; C:focal adhesion; IDA.  
 CC InterPro; IPR000299; Band 4.1.  
 CC InterPro; IPR002558; ILWEQ.  
 CC Pfam; PF00373; Band 41; 1.  
 CC Pfam; PF01608; I\_LWEQ; 1.  
 CC ProDom; PD011820; ILWEQ; 1.  
 CC SMART; SM00295; B41; 1.  
 CC SMART; SM00307; ILWEQ; 1.  
 CC PROSITE; PS00660; FERM\_1; 1.  
 CC PROSITE; PS00661; FERM\_2; 1.  
 CC PROSITE; PS00057; FERM\_3; 1.  
 CC PROSITE; PS00945; I\_LWEQ; 1.  
 CC PROSITE; PS00945; I\_LWEQ; 1.  
 KW Structural protein; Cytoskeleton; Phosphorylation.  
 FT DOMAIN 86 403  
 FT FERM.  
 FT DOMAIN 2340 2533  
 FT VARIANT 1105 1105 L -> P.  
 FT VARIANT 2180 2180 K -> M.  
 SQ SEQUENCE 2541 AA; 269832 MW; 14EF75ABE9FC2CBB CRC64;  
 Query Match 2.3%; Score 219.5; DB 1; Length 2541;  
 Best Local Similarity 19.4%; Pred. No. 0.025;  
 Matches 444; Conservative 295; Mismatches 860; Indels 693; Gaps 110;  
 QY 11 KAAVHTAANPVG---HG-VALQ---QGSSSSSPQN-----AAASLAEEGNKRMKMPR 56  
 DB 428 KSTVLQQYNNRVGKVEHGSVALPAMRSAGSGPENFQVGSMPPAQQOITSGQMRGHMPP 487  
 QY 57 IHQPSTAADG-----ISAAHQKKFSRLRCLG-----TKFPSAPQGPQTTHSK- 103  
 DB 488 LISAQAALGTGTINSSWQAVQAQATLDDFTLPLGQDAASKAWRNKNKMDSEKHEIHSQV 547  
 QY 104 -----GATLRDLLARDGGETOHE-----AAAPDAARLTR---SGGVK---RRNMDMA---GRP 148  
 DB 548 DAITAGTASVNLTAGDPAETDVTAVGCAVTTISNLTMSRGVKKLLAALLEDEGGNGRP 607  
 QY 149 MVKGGSGEDKVPFQQRHQLNFGQMRQTMLSKVAHPASNAGDRLOH-----SPHPIDP 203  
 DB 608 LQQAARGLAGAVSELLRSQAASAPRQNLQAGNVGQA---SGELLQOIGESDTPDFHQD 666  
 QY 204 SHEIKEEPPVGSTSKATAHADRVETAEQDDSEFQQLHQRLARERENPPQPKLGVA 263  
 DB 667 VLMQL-ANAVASAAALVLKAK---SVAQRTEDSLQ---TQVIAAATQCALSTSLVACT 720  
 QY 264 PISARFQPKLTA-----VAESVLEGTDTTQSPKLPQSMKMGSGAGVTPLAY 310  
 DB 721 KVVA---PTISSPVCOQLVEAGRLNAKAVEGCVSQAATEDGQLLRGVGAAT--AVT 775  
 QY 311 LDKGKLQALPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKCHLFDIKSTAT 370

Db 776 -----QALNELL-----CHVKAHATGAGPAGR-----YD-QATDT 804  
 QY 371 SYSVLHN-----SHPGEIKGK---LAQAGT---GSYSVDGSKGKISLGSGTQSHNKTMLS 419  
 DB 805 ILTVTNIPISSMGDAGEMVRQARILQAUTDLVNAIKADAE-----GESDLENSKILTS 858  
 QY 420 QP-----GEAHSRLTIGIWOHPAGAAAPQGESIRLHDDKIHLHPGLGVMQASDKDTHSLS 476  
 DB 859 AAKILADATAKMV-----EAAKGAA-----AHPD-----SEEQOQLR 891  
 QY 477 ROADGKLYALKONRTYQNLSDNKSEKLVYDIKYSVDQRGQVAILTDTPRHMSIMPS 536  
 DB 892 EAAEG--LRMATNAAQ-----NAIKKLVQRL-  
 QY 537 LDASPESHISLSLH-PADAHQGLHCKSELEAQSVASHGRLLVVADE-----GKLFSA 589  
 DB 944 ASAGPQLLVQSCKVAEQIPLLVQVGRSQAPDSFSAQLALIAASQSFQGGKRWAA 1003  
 QY 590 A-----IPKQDGNELKMKAMPQHALDEHFGHDHIOISGFFHDDHGLNALVKN----- 637  
 DB 1004 AKASVPTIQD-----QASAM-----QLSQCAKNLGTALAE 1035  
 QY 638 --NFRQOHAC-PLGNHDHQPFGWNL-----TDALVIDNLGLHHTNPPEHIL-----D 683  
 DB 1036 TRAQKAQEAACGPLEMDSALSVVQNLKDLQEIKAARADGKL-----KPLPGETMEKCTQD 1090  
 QY 684 MGH-----LGSALQEGKHLHYFDLTGKWTGAESDCQKLGKGLDGAAYLLKDG 731  
 DB 1091 LGNSTKAVSSATAKLGEIA--QGNENYAGIAROVAGG---LRSIAQAARGVAALTS 1145  
 QY 732 EVKRLINOSTSIRKHGTENVFSLPHVRNKPEPGDALQGLNKDDKAQAMAVIGNKYL-- 789  
 DB 1146 AVQAIVLDTASDVLDDKASSLIBEAKKASGHPGPEQSRQAQVAKAVTQA---LNRVC 1202  
 QY 790 -----ALTEKGD-----IRSFQIKPTQQLERPAQILSRGIGS-----ELKDIHV 830  
 DB 1203 LPQQRDVNLRVAVGDAKRLSLDLPLPSTGTGTQE-AQSRLENEAAGLNAQATLVQASR 1261  
 QY 831 DHKQNL-YALTHEGEVHFQPREAWQNGAESSSWHKALPQSE-----SKLSLDMSEH 883  
 DB 1262 GTPQDLARASGRFGQDFSTFLEA---GVEMAG---QAFQEDRAQVSVNLKISMS 1314  
 QY 884 -----KPIATFEDGSQHLKAGGWHAAYAAPERGLAVGTSGSQTV----- 923  
 DB 1315 LLLAAKALST--DPASPFLKS---QLAAAAA---AVTDSINQLITMCTQAPQKQEC 1365  
 QY 924 -----FNRLMQGVKGVIPGSGLT-VKLSAQTG-----GMT 953  
 DB 1366 ALFQLETVRELLNPVQPIINDMSYFGCLDSVMENSKVLGEAMTGISQNAKGNLPRFGDA 1425  
 QY 954 GABGRKVSFKSERIRAYAFNPTMTPRPIKNAAYATQHWQREGKLPKLYEMOGALIKQ 1013  
 DB 1426 IATASALCGFTEAAQAAAYLVGVSDP-----NSQACQGLVPTQARA--- 1470  
 QY 1014 LDHNVHRNAPQDLQSLKLETLDLGBHG---ABELNDMKRFRDELQBSATRSVTVLGQH 1070  
 DB 1471 -----NQAIQMACQS-----LGPFGCTQAOVLS-----AATIVAKHT 1502  
 QY 1071 GVILKSNGEINS--EFKPSFGKALVQSF-VNNSRGDLSKSLQ-----QAVHAT 1115  
 DB 1503 SALCNSCRLASARTANTPTAKRFQVSAKEVANSTANLVTKIKALDGDGFTTEENRAQ 1562  
 QY 1116 PPSAES--KLQSMGLH-----FVSJAGDMSHQKGEIPLGR--- 1148  
 DB 1563 APLEAVDNLASFASPESSVPAQISPEGRAAMEPISAKTMLESAGSLIQTARALAV 1622  
 QY 1149 -QRDPNDKALT-KSLIILDTV----- 1168  
 DB 1623 NRPDPWRWSVLGHSRTVSJSDSKLITSMRDKAPQLECEITAALNSCLRLDQASLAA 1682  
 QY 1169 -----TIGELHELADKAKLVSDHKPDADQ---IKOLRQOFD- 1201

Db 1683 VSQALPREGISQBALHTQMTAVQEISHLIE--PLASAAAEAEASQLGHKVKVQAQYFEP 1740  
Qy 1202 -TLREKREYESPVKHYTDMGETH-----NKALEANYDAVKAFAFNAFKKEHGHVNLNLT 1254  
Db 1741 LTLAAVGAASKTSLHPQOMALLDQTKTLEASALQLLTYTAKAGNPKAAH-----TQE 1794  
Qy 1255 VLSE--QGSAAELAKLNTLSLDSGESMSFSRSGYGGVSTVFVPTLSKKVPVPIQAG 1312  
Db 1795 ALEBAVQMTAEVDELATTL-----NEAASAAGVVGGMVDSI-TQAINQLDGGPMGDEG 1848  
Qy 1313 ITLDRAVLSFRTSGGLNVSF-----GRDGGVSGNIMVATGHDVMPYTGKKS 1362  
Db 1849 SFVD--YQTTMVRTAKAVTAVQEMVTKNSPBEPLANQLTSDYGRLASQAKPAVA 1906  
Qy 1363 AGNASDMSAKHKISPDILIRIGAAVSGTQGTQLQNSLKFELTEDELPGFIHGLTHGTLTPA 1422  
Db 1907 AENEIEGAHKKRVO-ELGHCCSALVTKAGALQCSPSDVYTKEL-----TECA 1954  
Qy 1423 ELLOKQIEH---QMKQSKLTFSVDTSAN-----LDLRAGI-----NLNEDGSKP--- 1464  
Db 1955 RRVSEKSVHLAALQAGNRGTQACITTAASAVSGIADLDTTIMEATAGTLNREGAETPAD 2014  
Qy 1465 --NGV--TARVSA-GLSASANLAGSRERSTTSGQ-----FGSTTSASNNR 1505  
Db 2015 HREGILTKAVLVEDTKVLQVNAAGSQEKLAAQASQSVATITRLADVVKVGAALGADP 2074  
Qy 1506 PTFNLGVGAGANLTAALGVASHSTH--EGKPVGIFPAFTSTNVSAALALDNRTS----- 1557  
Db 2075 ETQVVLINAVKAVAKGLDLISATKAAAGK-VGDDPAVWQNLKSAKVMVNTVTSLLTKVK 2133  
Qy 1558 -----QSTISLEL-----KRAEPTVTSNDISBLTSTLGHFKDSATTK 1593  
Db 2134 AVEDEATKGTALREATTETHEIRQELAVFCSPPEPAKTSPTDFIRMTKGI-----TMATAK 2188  
Qy 1594 MLA-----ALXELDKAEQHLILQHFSAKDVVGDYERAEVNRMLKLIV----- 1638  
Db 2189 AVAAGNSCROEDVITANLSRAADMLRAACKAEAFHPEVADPVRLRALHYGREANGYL 2248  
Qy 1639 -----IRQAAADSHSMELGSASH-----STTYNNLS 1664  
Db 2249 ELLDHVLLTLQKPNPDLKQLTGHSKRVAGSVTELIQAAEAMKTEWDPEDPTVIAENE 2308  
Qy 1665 RINNDGIVELHKLHFDALPASSAKRIGEMNNPDALKDIIKQISTPFSSASVSMELKD 1724  
Db 2309 LLGAAAAIEAAAKLEQLKPRAPKEADESLNFEQILEAAKSI-----AAATSALVKA 2362  
Qy 1725 GLRQTEKAILDGVKREVGVLFDQRNLRVKSVSQSVKS-----EGFN 1772  
Db 2363 ASAAQRE-LVAQGVKAIPANAL--DDGQWSQGLISAARMVAAATNNLCEAANAQVQHA 2419  
Qy 1773 TPALLLGTSAAAMSMERNIGTINFKYQDQNTPRFTLEG-GIAQAN-----PQVASA 1825  
Db 2420 SQEKLISAKQVAASTAQLLVACKVKADQDSEAMKRLQAAGNAVKRASDNLVKAQKAA 2479  
Qy 1826 LTDLKKEGLEMK 1837  
Db 2480 FEDQENETVVVK 2491

RESULT 7  
MLL2 HUMAN  
ID MLL2 HUMAN  
AC O14686; O14687;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein).  
DE MLL2 OR ALR.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RX MEDLINE=97388474; PubMed=9247308;  
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,  
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,  
RA Canaani E.;  
RT "Structure and expression pattern of human ALR, a novel gene with  
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila  
RT trithorax.";  
RL Oncogene 15:549-560(1997).  
RN [2]  
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.  
RC TISSUE=Cervical carcinoma;  
RX MEDLINE=22371496; PubMed=12482968;  
RA Goo Y.-H., Sohn Y.C., Kim D.-H., Kim S.-W., Kang M.-J., Jung D.-J.,  
RA Kwak E., Barlev N.A., Berger S.L., Chow V.T., Roeder R.G.,  
RA Azorsa D.O., Meltzer P.S., Suh P.-G., Song E.J., Lee K.-J., Lee Y.C.,  
RA Lee J.W.;  
RT "Activating signal cointegrator 2 belongs to a novel steady-state  
RT complex that contains a subset of trithorax group proteins.";  
RL Mol. Cell. Biol. 23:140-149(2003).  
CC -!- FUNCTION: May be involved in transcriptional regulation.  
CC -!- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which  
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/  
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins  
CC MLL2 and MLL3, and ASH2/ASCL2.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Name=1;  
CC IsoId=O14686-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=O14686-2; Sequence=VSP\_008563, VSP\_008559;  
CC Name=3;  
CC IsoId=O14686-3; Sequence=VSP\_008560;  
CC -!- TISSUE SPECIFICITY: Expressed in most adult tissues, including a  
CC variety of hematopoietic cells, with the exception of the liver.  
CC -!- MISCELLANEOUS: This gene mapped to a chromosomal region involved  
CC in duplications and translocations associated with cancer.  
CC -!- SIMILARITY: Belongs to the transcription factor trithorax family.  
CC -!- SIMILARITY: Contains 5 PHD-type zinc fingers.  
CC -!- SIMILARITY: Contains 1 post-SET domain.  
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.  
CC -!- SIMILARITY: Contains 1 SET domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF010403; AAC51734.1; -;  
CC EMBL; AF010404; AAC51735.1; -;  
CC PIR; T03454; T03454.  
CC PIR; T03455; T03455.  
CC Genew; HGNC:7133; MLL2.  
CC MIM; 602113; -;  
CC GO; GO:0005634; C:nucleus; TAS.  
CC GO; GO:0003700; F:transcription factor activity; TAS.  
CC GO; GO:0007048; P:oncogenesis; TAS.  
CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.  
CC InterPro; IPR003889; FYrich\_C.  
CC InterPro; IPR003888; FYrich\_N.  
CC InterPro; IPR000910; HMG\_12\_box.  
CC InterPro; IPR003616; PostSET.  
CC InterPro; IPR006118; Recombinase.  
CC InterPro; IPR001214; SET.  
CC InterPro; IPR001965; Znf\_PHD.  
CC InterPro; IPR001841; Znf\_ring.  
CC Pfam; PF00628; PHD; 5.  
CC Pfam; PF00856; SET; 1.









Db 1472 -----KSFELASAVEKLGALQKLHSELSL-----MENIKSQLEAKEKIQV 1513  
 Qy 1588 DSATTKMLAALKEADDAPKPAQLHLQOHFSAKDVGVDERVEAVRNKLVIRQQAADSH 1647  
 Db 1514 DESP-----IQEJLHEITASK-----NNYEGKLNKDSIIRDLSS----- 1547  
 Qy 1648 SMELGSASHSTYNNLRINNDGIVELHKKHFDALPASSAKRLGEGMMNNDPALKDIIKQ 1707  
 Db 1548 -----ENIEQLNN-----LL-----AEKSAVKRLSTE-----KE 1572  
 Qy 1708 LQSTPPFSASVSMELKQGLREQTBAKILDKGVKEEVGVLPQDRNNLRKVSQSVSK 1767  
 Db 1573 SEILQFNSRLADLEY-----HKSQVESBLG-----RSKLKLASTTBELQLAE 1614  
 Qy 1768 SEGNTFALLIGTNSAAMSMERNIGTFKYGQDQNTPRFTLEGGAQANPQ----- 1821  
 Db 1615 NERLSLTRMLDLQNV-----KOLSNIKOSLSLSEDLRTL-----SLEDSVASLQKECKIKSN 1667  
 Qy 1822 ----VASALTDLKKEGLEMK 1837  
 Db 1668 TVESLQDVLTISVQARNAELE 1687

RESULT 9

AK12 HUMAN  
 ID AK12 HUMAN STANDARD; PRT: 1781 AA.  
 AC Q02952; Q00310; Q00498; Q99970;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE A-kinase anchor protein 12 (A-kinase anchor protein 250 kDa) (AKAP  
 DE 250) (Myasthenia gravis autoantigen gravin).  
 GN AKAP12 OR AKAP250  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Heart;  
 RX MEDLINE=97153077; PubMed=9000000;  
 RA Nautert J.B., Klauck T.M., Langeberg L.K., Scott J.D.;  
 RT "Gravin, an autoantigen recognized by serum from myasthenia gravis  
 RT patients, is a kinase scaffold protein.";  
 RL Curr. Biol. 7:52-62(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Unbilical vein endothelial cells;  
 RX MEDLINE=9269042; PubMed=9604001;  
 RA Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;  
 RT "Changes of gene expression by lysophosphatidylcholine in vascular  
 RT endothelial cells: 12 up-regulated distinct genes including 5 cell  
 RT growth-related, 3 thrombosis-related, and 4 others.";  
 RL J. Biochem. 123:1119-1126(1998).  
 RN [3]  
 RP SEQUENCE OF 43-1781 FROM N.A.  
 RC TISSUE=Unbilical vein endothelial cells;  
 RA Bowditch R.D., Ginsberg M.H.;  
 RT "Sequence of gravin cDNA isolated from a human umbilical vein  
 RT endothelial cell library.";  
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 1477-1781 FROM N.A.  
 RC TISSUE=Unbilical vein endothelial cells;  
 RX MEDLINE=92395179; PubMed=1522245;  
 RA Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,  
 RA Lindstrom J., Ginsberg M.H.;  
 RT "Molecular cloning and preliminary characterization of a novel  
 RT cytoplasmic antigen recognized by myasthenia gravis sera.";  
 RL J. Clin. Invest. 90:992-999(1992).  
 CC -!- FUNCTION: Anchoring protein that mediates the subcellular  
 CC compartmentation of protein kinase A (PKA) and protein kinase C

(PKC).  
 -!- SUBUNIT: Binds to dimeric RII-alpha regulatory subunit of PKC.  
 -!- SUBCELLULAR LOCATION: Cytoplasmic. May be part of the cortical  
 cytoskeleton.  
 -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q02952-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q02952-2; Sequence=VSP\_004110, VSP\_004111;  
 CC Note=No experimental confirmation available;  
 CC TISSUE SPECIFICITY: Expressed in endothelial cells, cultured  
 CC fibroblasts and osteosarcoma, but not in platelets, leukocytes,  
 CC monocytic cell lines or peripheral blood cells.  
 CC -!- INDUCTION: Activated by lysophosphatidylcholine (lysopC).  
 CC -!- DOMAIN: Polybasic regions located between residues 265 and 556 are  
 CC involved in binding PKC.  
 CC DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY  
 CC PATIENTS WITH MYASTHENIA GRAVIS (MG).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; U81607; AAC51366.1; -.  
 CC EMBL; AF001504; AAB58938.1; -.  
 CC EMBL; AB003476; BAA19927.1; -.  
 CC EMBL; M96322; AAA35931.1; -.  
 CC Genew; HGNC:370; AKAP12.  
 CC MIM; 604598; -.  
 CC DR GO; GO:0005737; C:cytoplasm; TAS.  
 CC DR GO; GO:0005079; F:protein kinase A anchoring activity; TAS.  
 CC DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; TAS.  
 CC DR InterPro; IPR001573; PKinA anch.  
 CC DR Pfam; PF03832; PKinA anch.  
 CC DR Antigen; Repeat; Alternative splicing.  
 CC FT DOMAIN 603 633 AKAP 1.  
 CC FT DOMAIN 752 782 AKAP 2.  
 CC FT DOMAIN 797 827 AKAP 3.  
 CC FT DOMAIN 98 101 POLY-GLU.  
 CC FT DOMAIN 265 556 INVOLVED IN PKC-BINDING (PROBABLE).  
 CC FT DOMAIN 1540 1553 RII-BINDING (PROBABLE).  
 CC FT VARSPLIC 1 98 Missing (in isoform 2).  
 CC FT VARSPLIC 99 106 /FTId=VSP\_004110.  
 CC FT VARSPLIC 117 117 EEEVITE-> MLGTTIT (in isoform 2).  
 CC FT CONFLICT 117 117 /FTId=VSP\_004111.  
 CC FT CONFLICT 142 144 E-> K (IN REF. 2).  
 CC FT CONFLICT 215 215 NRN -> TPEI (IN REF. 2 AND 3).  
 CC FT CONFLICT 448 448 Q-> K (IN REF. 2).  
 CC FT CONFLICT 694 694 G-> E (IN REF. 2 AND 3).  
 CC FT CONFLICT 867 867 R-> G (IN REF. 2 AND 3).  
 CC FT CONFLICT 867 867 G-> S (IN REF. 2 AND 3).  
 CC FT CONFLICT 986 986 S-> A (IN REF. 2 AND 3).  
 CC FT CONFLICT 1530 1530 E-> EE (IN REF. 3 AND 4).  
 CC FT CONFLICT 1581 1581 V-> M (IN REF. 4).  
 CC FT CONFLICT 1601 1601 Q-> L (IN REF. 2).  
 CC SQ SEQUENCE 1781 AA; 191439 MW; BA813937379FAC0F CRC64;  
 CC -----  
 CC Query Match 2.3%; Score 214; DB 1; Length 1781;  
 CC Best Local Similarity 17.9%; Pred.No. 0.026;  
 CC Matches 339; Conservative 278; Mismatches 718; Indels 564; Gaps 80;  
 CC -----  
 CC Qy 29 QGSSSSSPQAAALAAEGKNGKMPRIHQPSAADG-----ISAHQOKKSFSLRGCL 83  
 CC Db 245 EQSHAEISPAESQVAECKEKEGKEQKESAPSPVTSFGTSFKFTQGWA 304  
 CC Qy 84 GTKK---FSRSAPQGQGTTHSKATLURLDARDGDTQHEAAAPDAARLTRSGGVKRN 140  
 CC Db 305 GWRKTSFRPKPEDEVEASEKKKEQEKVDTEEDGK-----ARVASEKLTAASQAHPQE 359



```

RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 0509952;
EX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sakakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22 (2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: Belongs to the intimin/invasin family.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB005423; AAG57041.1; -
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big_1;
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR008964; Invasin_intimin.
DR Pfam; PF02369; Big_1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID_1; 16.
DR SMART; SM00889; PKD; 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834 BIG-1 1.
FT DOMAIN 840 929 BIG-1 2.
FT DOMAIN 931 1033 BIG-1 3.
FT DOMAIN 1042 1132 BIG-1 4.
FT DOMAIN 1134 1236 BIG-1 5.
FT DOMAIN 1245 1335 BIG-1 6.
FT DOMAIN 1337 1439 BIG-1 7.
FT DOMAIN 1448 1539 BIG-1 8.
FT DOMAIN 1548 1652 BIG-1 9.
FT DOMAIN 1653 1750 BIG-1 10.
FT DOMAIN 1751 1855 BIG-1 11.
FT DOMAIN 1856 1957 BIG-1 12.
FT DOMAIN 1963 2056 BIG-1 13.
FT DOMAIN 2065 2156 BIG-1 14.
FT DOMAIN 2157 2252 BIG-1 15.
FT DOMAIN 2254 2355 BIG-1 16.
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;

Query Match
Best Local Similarity 2.2%; Score 205; DB 1; Length 2660;
Matches 379; Conservative 246; Mismatches 775; Indels 578; Gaps 92;

QY 170 NFGQMRQTML-----SKMAHPASANAGDRLOHSPHPPHSGSHHEIKKEEPVGTGSK 218
DB 514 NFNREQSMVVVQAPTLSQKDSVSLSSQTLSD--SHSTATLTFTIAHDAAGNEPVLG-V 570
QY 219 ATTAHADRVEIAQD-----DDSEFOQLHQQLARERENPPQ-----PPKLGVATP 264
DB 571 LSTRHEGVQDITLSDWKDNGSGVTQILTTGAMSGTILTPQLNGVDAKAPAVNNIISV 630
QY 265 ISARFQPKLTAVARSVLEGT-----DTQSPKLPQSMKSGAGVTPPLATLDKKG 315
DB 631 SSSRTHSSIKDKRYLSGNPIEVTVELRDENDKPVKEQKQQLAT-----AVSIDNVK 683
QY 316 LQLAPDNPALNTLLKQTL-OKOTQHYLAHASSDGGQHLLLDN-----KGLHFDIKS 367

```

```

DB 684 -----PGVTTDWKETADGVYKATYATYTKGSLTAKLLMQNWNEDLHTAGFIIDANP 735
QY 368 TATSYSVLHNSHPCEIKKLAQAAGTGSVSV--DGKSGKIS-----LGSGTQSHN--- 414
DB 736 QSAKIATLSASNGNGLANENA-ANTVSVNVADEGSPNDHTVTFVAVLSSGATSFNQNT 794
QY 415 -KTMLSOPGEAHRSLLTGIMQHPAGAARPOGESIRLHDDKIHLHPGLGVWQSA----- 467
DB 795 AKTDVN-----GLATFDLKSSKQEDNTVEVT-LENGVQKTLIVSVFG 835
QY 468 -----DKDTHSQLSRQAQKLYALKDNRTLQMLSNKSEKLUVD 506
DB 836 DSSTAQVDLQSKNEVVADGNSATMTATVRDAKGNL--LNDVKVTFNVNAAA----- 887
QY 507 KIKSYVDQRQVAILLTDTPGRH-KMSIMPSLDASPESHISLSLHFADAHQGLLHGKSEL 565
DB 888 KLSQTEVNSHDGIATATLSLKNGDYVTASVSSGQAN-----QQVIFIGDOST 937
QY 566 EAQSVAIHSHGRVLVADSEGLFSAAI PKQDGNELKMKAMPQHALDEHFGHDHQTSGFFH 625
DB 938 AALTLSVPSGDITVTNT-APLHMTATLQDKNGNPLKKEIT-----FSVPNDVASRFS 989
QY 626 -----DDHGQLNALVKNNFRQOH--ACPLGNHQHFHPCWNLTD-----ALVIDNQLG 670
DB 990 ISNSGKGMTDSNGTAIASLTGTLAGTHTMITARLANS-----NVSDTPQMTFFVADKDA 1042
QY 671 LHHTNPEPHEITLDMHGLSLALQGLKHLHYFDQLTKGWTG----- 709
DB 1043 VVVLQTSKAEIINGG-VDETTLTATVKDPPDNVNVKNLSVVPRTSPADTQLSLNARNNEN 1101
QY 710 --AESDCKQLKGLDGAAYLLKDG--EYKRLININQSTSS-----IKHGTENVF 753
DB 1102 GIAEVTLKGTVLGVHTAEAILLNGNRDITKI VNIAPDASNAQVTLNI PAQVVTNNSDSVQ 1161
QY 754 SLPHVR--NKPEPGDALQGLNKDDKA-----QAMAVIGVNKYLAITEKGDIRSPOIK 803
DB 1162 LTATVKDPSNHPVAGITVNFPTMPQDVAANFTLENNGIAITQANGEAHVTLKGG--K 1215
QY 804 PGTQOLERPACTLSREGISGELKDIHVHDKQNLVAL--THEGEVPHQPREAWQNGAES 861
DB 1216 AGHTTV---TATLGNNNASDAQVTVFVADKDSAVVVLQTSKAEI-----GNGVDETT 1265
QY 862 WHKLALPQSESKLKLSDMSHEHKPIATFEDGSQHLKAGG-----WHAAAP 908
DB 1266 LTATVKDPPDNVAVKDLQVTFSTNPADTQLSQSKNTNDSGVAEVTFKGTVLGVHTAEATL 1325
QY 909 ERG-----PLAVGTSGSQTVFNRLMQGVKVIPOGSLTVKLKSAQTGGMTGAEGRKVS 961
DB 1326 PNGNNDTKI VNIAPDASNAQVTLNIPAQ-----QVVTNNSDSVQLTA-----TVK 1370
QY 962 SKFSERTRAVAFNPTMTSTPRPI-----KNAAVATQHGMCQREGKLPXYEMOGALIKQL 1014
DB 1371 DPSNHPVAGITVNFMT--PDQVAANFTLENNGIAITQANGEAHVTLKGGKAGTHTVTATL 1428
QY 1015 DAHVNRHNAFPQDLQSKLETIDLGEHGAELNDMKRFRDELE-----QSATRSVTVLGQHQ 1070
DB 1429 SNNNTSDSQPVTFFVADKTSALV-----LQISKNEITGNGVDSATLITATVKDQF- 1477
QY 1071 GVUKSNGEINSEFKPSGPKALVQSFNVNRSGDLSKSLQQA VHAATPPSABSKLSQMLGHF 1130
DB 1478 -----DNEVNN-----LPVTFSTASSGLTIT-----PESNTNESGIAQA 1512
QY 1131 VSAGVDMSHQGEIPLGRQRDNDKATLTKSRLITDTVTIGELHELADKAKL-VSDHKPD 1189
DB 1513 TLAGVAFGEQTVTASLANNGASDNKT-----VHFIGDTAAAKIETLTPV 1556
QY 1190 ADQI---KQLRQOQFDTLREKRYES--PVKHYTDMGFTHNKALEANYADVAKFINAFKKE 1244
DB 1557 PDSIIAGTPQNSGSSVITATVVDNNGFPVKGVTVNFVTSNAAATAEIMTNGGCAVTN-----E 1611
QY 1245 HHGVNLT---TRTVLES-----QGSFAELAKKLK-----NTLL-----SL 1275
DB 1612 QGKATVYTTNTRSSIESGARPDVTEASLENGSSITLSTSINVNADASTAHLTLLQALPDTV 1671

```



CC IsoId=P55937-2; Sequence=VSP\_007730;  
CC Note-No experimental confirmation available;  
CC TISSUE SPECIFICITY: Highly expressed in testis. Transcripts can be  
CC found in spermatids during spermatogenesis. No expression in  
CC Leydig cells, spermatogonia or spermatocytes. Detected at low  
CC levels in all tissues.  
CC -!- DOMAIN: Extended rod-like protein with coiled-coil domains.  
CC -!- PTM: Cleaved by caspases in apoptotic cells (By similarity).  
CC -!- PTM: Phosphorylated (By similarity).  
CC -!- SIMILARITY: Belongs to the golgin family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; D78270; BAA19612.2; -.  
CC EMBL; AB029537; BAA86889.2; -.  
CC EMBL; AB029521; BAA86889.2; JOINED.  
CC EMBL; AB029522; BAA86889.2; JOINED.  
CC EMBL; AB029523; BAA86889.2; JOINED.  
CC EMBL; AB029524; BAA86889.2; JOINED.  
CC EMBL; AB029525; BAA86889.2; JOINED.  
CC EMBL; AB029526; BAA86889.2; JOINED.  
CC EMBL; AB029527; BAA86889.2; JOINED.  
CC EMBL; AB029528; BAA86889.2; JOINED.  
CC EMBL; AB029529; BAA86889.2; JOINED.  
CC EMBL; AB029530; BAA86889.2; JOINED.  
CC EMBL; AB029531; BAA86889.2; JOINED.  
CC EMBL; AB029532; BAA86889.2; JOINED.  
CC EMBL; AB029533; BAA86889.2; JOINED.  
CC EMBL; AB029534; BAA86889.2; JOINED.  
CC EMBL; AB029535; BAA86889.2; JOINED.  
CC EMBL; AB029536; BAA86889.2; JOINED.  
CC EMBL; AB029537; BAA86889.2; JOINED.  
CC EMBL; AB029521; BAA86890.2; -.  
CC EMBL; AB029522; BAA86890.2; JOINED.  
CC EMBL; AB029523; BAA86890.2; JOINED.  
CC EMBL; AB029524; BAA86890.2; JOINED.  
CC EMBL; AB029525; BAA86890.2; JOINED.  
CC EMBL; AB029526; BAA86890.2; JOINED.  
CC EMBL; AB029527; BAA86890.2; JOINED.  
CC EMBL; AB029528; BAA86890.2; JOINED.  
CC EMBL; AB029529; BAA86890.2; JOINED.  
CC EMBL; AB029530; BAA86890.2; JOINED.  
CC EMBL; AB029531; BAA86890.2; JOINED.  
CC EMBL; AB029532; BAA86890.2; JOINED.  
CC EMBL; AB029533; BAA86890.2; JOINED.  
CC EMBL; AB029534; BAA86890.2; JOINED.  
CC EMBL; AB029535; BAA86890.2; JOINED.  
CC EMBL; AB029536; BAA86890.2; JOINED.  
CC EMBL; BC043452; AAH3452.1; -.  
CC EMBL; BC053002; AAH33002.1; -.  
CC EMBL; AK032610; BAC27949.1; -.  
CC PIR; T42722; T42722.  
CC MGD; MGI:96958; Golga3.  
CC GO; GO:0005793; C:ER-Golgi intermediate compartment; IDA.  
CC GO; GO:0000139; C:Golgi membrane; IDA.  
CC GO; GO:0005515; F:protein binding; IPI.  
CC InterPro; IPR003345; M\_repeat.  
CC Pfam; PF023370; M: 5.  
CC Antigen; Spermatogenesis; Developmental protein; Golgi stack;  
CC Coiled coil; Phosphorylation; Alternative splicing.  
CC FT DOMAIN 132 217 GOLGI-TARGETING DOMAIN (BY SIMILARITY).  
CC FT DOMAIN 198 281 SER-RICH.  
CC FT DOMAIN 323 326 POLY-ALA.  
CC FT DOMAIN 318 1414 COILED COIL (POTENTIAL).  
CC FT DOMAIN 494 694 GLN-RICH.  
CC FT DOMAIN 1187 1325 GLN-RICH.  
CC FT VARSPLIC 96 96 G -> GVDGFHDNLRNSQGTSAEGSVERKEALQSLRLSLPMQ

FT FT CONFLICT 30 30 ETOLC (in isoform 2).  
FT FT CONFLICT 437 437 /FTId=VSP\_007730.  
FT FT CONFLICT 552 552 T -> I (IN REF. 3; AAH53002).  
FT FT CONFLICT 803 803 G -> E (IN REF. 3; AAH53002).  
FT FT CONFLICT 803 803 M -> T (IN REF. 2).  
FT FT CONFLICT 942 942 Q -> L (IN REF. 2).  
FT FT CONFLICT 975 975 T -> S (IN REF. 3; AAH53002).  
FT FT CONFLICT 1426 1426 A -> S (IN REF. 3; AAH43452).  
FT FT CONFLICT 1447 AA; 162820 MW; 88B68F2D142F8A9B CRC64;  
SQ SEQUENCE 1447 AA; 162820 MW; 88B68F2D142F8A9B CRC64;  
  
Query Match 2.2%; Score 204; DB 1; Length 1447;  
Best Local Similarity 18.9%; Pred. No. 0.057;  
Matches 316; Conservative 257; Mismatches 675; Indels 426; Gaps 77;  
  
QY 112 ARDQ-----ETQHEAAAAPDAARLTRSGGVKRNMDMDAGRPMVKGGSG--EDKVPTQOK 164  
DB 6 AKODGLWESKSSDDVSCPEASLETVGSRLARLPDQDTAQDASVENVNRGKSGSPDRSS 65  
  
QY 165 RHQNNFGQMRQTMLSKMAHPASANAGDRLQHSPPHIPGSHHEIKEEPPVGSSTKATTAHA 224  
DB 66 QVAICQNGQIPDLQLS--LDPTTSPVG-----PDASTGSTASSLPLEK 106  
  
QY 225 DRVEIAQEDDDSEFQQLHQORLARERENPPPKLGVATPISARFQPKLTAVAESVLEGT 284  
DB 107 EQVRLQARKLE-EQLMQYRVKRHERSSQP--ATQMKLFSTLDPELM-----152  
  
QY 285 DTTQSPKPKQSMKLGSGAGVTPLAVTLDKGLQLAPDNPPALNTL-----LKQT 333  
DB 153 -----LNFENLPRAS-----TVAVTKEYSFLTSVPRGPKVSLGSLLAHSKKKSS 201  
  
QY 334 LKQDTQHYLAHASSDGSQHLLDNKGHLFDIKSTATSYSLVNSHPGEIKGKLAQAGT 393  
DB 202 KIRSLADYRTEDPSDGLGSTADAVGS--SLQSRSSSTSVSEVSPSSTDNKRV-----255  
  
QY 394 SVSVGKSKGKISLGSTQSHNKTMLS-----QPGEAHRSLITG-----IWQ 434  
DB 256 SASMTGDSVSEADGNESDSSSHSLSARGACGVLGNGVMPGTAY--MVDQETSAEALGQ 313  
  
QY 435 HP-----AGAAPQGES-----IRLHDDKTHILPELVGWSADKQTHSOLSRQADG 481  
DB 314 FPSIKQVLOQAAAHQHQOQNEANGEVRSRDSI-----CSSVSMESSLAEPODE 362  
  
QY 482 KLYALKDNRTLQNLSDN---KSSEKLVDKIKSYSDVQRGQVAILT---DTPGRHKMSIMP 535  
DB 363 LLQILKDKRLEQGEVEALSLEASQALEK-----AELQAQALALSTRLOAQVEHSHSQ 417  
  
QY 536 SLDASPESHISLHPADAHQGLHKGSELEAQ--SVATSHGLRVADSEGLKFLSAIPK 593  
DB 418 KQDSLSEVDTLKQSCWDLGRAMTDLQSMLEAKNASLASNNDLQVAEEQYQRLMA----473  
  
QY 594 QGDGNEKMKAMPQHALDEHFGHDHOSIGFFHDDHQLNALVKNFNQHQACPLGNHDHF 653  
DB 474 -----KVEDMORNILSK-----DNIV-----HDLRQMTALQSL-----503  
  
QY 654 HPGWNLTDALVIDNQLGHLHNPPEPHEILDMHGLSLAQEGKLYFDQLTKMGTAESD 713  
DB 504 -----QQVQLERTLTISK--LQASQAEITSLQHAQWYQQQLT-----539  
  
QY 714 CKOLKGLDGAAYLLKDGVEKRLINQSTSS--IKH-GTENVFSLPH-----VRNK 761  
DB 540 -----LAQEARVRLQGEAHIQVQMTQAGLLEHLKLENV-SLSHQLTETQHRSIKEK 591  
  
QY 762 PEPGDALQGLNKDDKQAMAVIGVKNLYALTEKGDIRSFOIKPG-TQOLERPQATLSREG 820  
DB 592 ERIAVQLQSTEADMLDQEAAFVQIREAKTVESDLORRLEEFEGEREQKQVADAAA--S 649  
  
QY 821 ISGELDKIHV---DHKQNLVYALTHEG-EVPHQ---PREAWQNGAESSSWHKLALPOSEK 873  
DB 650 LEQQLQVXKLTLPORDQQLAALQOEHLVDVTKLTSTQEAALQAKQSLDLDLHTRYDELQAR 709  
  
QY 874 LKSLDMSHEHKPTAT-FEDGSGHQLKAGGHWAAPE---RGLPANGTSSQVTFNRLMQ 929  
DB 874 LKSLDMSHEHKPTAT-FEDGSGHQLKAGGHWAAPE---RGLPANGTSSQVTFNRLMQ 929

Db 710 LEELOREADSDRAHFLQNEKIVLEVALQSAKDEKELDRGARRLEEDTEET--SGLE 767  
Qy 930 GYVKGVIPGSLTVKLSAOTGWTGAEGKVSKEP-SERIRAVAFNPTWSTPRIPKNAAY 988  
Db 768 QIRODLAVKSNQVEHQQETATILR-KOMOKVKEQFVQKQWVEAYRDRATSKQDLINELK 826  
Qy 989 ATQHGQWG--REGLKPLYEMQG-----ALIKO-----LDAHVNRHNAPOF 1026  
Db 827 ATKRLDSEMKELRQBELIKQEGKKTVEHESFLQKDMSLVHQMAELEGHLSQVKERD 886  
Qy 1027 DLQSKLETLDLGHGABLND-----MKFRDELEQSATRSVTVLGHQGVLSKNGEINSE 1082  
Db 887 EMEIHLQSLKFKQEQMIALTEANETLKQIEELQEQAKAIT--EQQKQWKLUGSLDTA 944  
Qy 1083 FK--PSPGKALVQSFNVNRSQDLSKLOQAHVATPPSAESKLSQMLGHFVBSAGVD--MSH 1139  
Db 945 QKEMTKHKAYENAVSI-----LSRLQEL-ASKEATDAELNQLRAQSTGGSSDPVLH 997  
Qy 1140 QK-GEIPLGRQRPNDKTKALRSRLILDTVTIGELHELADKAKLVSDHDPADQIKQLRQ 1198  
Db 998 EKTRALEVELQNVGSKILLELQELQEVITWTSQELSEKREKVELED--ELQESRGFR 1054  
Qy 1199 QDTLREKYESPNVPHYDMGFTHNK--ALEANYDAVK-----AFINAPKKEHGV---- 1248  
Db 1055 KIKRUEES-----NKLALALEHERGKLTGLGQSNALRHNLSILETA 1097  
Qy 1249 -----NLTRTVLESQGSABLAKKLNTLLSDSGESMSFSRSGGVSTVFPV 1297  
Db 1098 LAKREADIVQNLQVAVL--QRKEEDROMQOLVOALQ-----VS 1136  
Qy 1298 TISKVVPVPIPGATILDRAYNLFSRTSGGLNVSFGRDGGVGNIMVATGHDVMPYMT 1357  
Db 1137 LEKERMEV-----NSLKEQMA 1152  
Qy 1358 GKTSAGNASDWLSAKHKISPDLRIGAAVSGTILQ-----TLQNSLKFKLTD-----E 1406  
Db 1153 ARIEAGH-----NRHFKAATLEL-SEVKELQAKEHLVQTQABVDELOQDKHSHOE 1206  
Qy 1407 LQGFTHGLTHGTLTPAELLQKIEHQWQ--GSK-----LTSVDTSAN--LDLRAGINL 1457  
Db 1207 IAQFQTELAEAR-TQQLLQKLEQMSQOPTGSEMEDLKWELQKEREISLQKQLODL 1265  
Qy 1458 NE-DGSKPQNGTVARVAGLSANLAAGSRERSTTSQF---GSTTSASNNRPTFLNGVG 1513  
Db 1266 TEQOQKKELEGTOQTTLTIKSELEMYQEDLS-ETQDKFELQAKVSELKNNKTLTQ--- 1321  
Qy 1514 AGANLTAALGVAHSSTHEGKPVGIFPAFTSNVSAALAL---DNRTSQSISLELKAEPV 1570  
Db 1322 --QNQOLKULDRRGAKKKEPKG-----BSNSSPATPIKIPCPVPASLLELLRPPPA 1374  
Qy 1571 TSND-ISELTSTIG--KHFKDSATTKW-----LAALKELDDAKPAEQH 1611  
Db 1375 VSKEPLKNLNCUQLKQEMDSIQRQWEEHTITVHESLSWAGV-EAAPAEHAH 1427  
RESULT 12  
ID APC MOUSE STANDARD; PRT; 2845 AA.  
AC Q61315; Q62044;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Adenomatous polyposis coli protein (APC protein) (mAPC).  
GN APC.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.  
RC STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;  
RX MEDLINE=92263101; PubMed=1350108;  
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,

RA Luongo C., Gould K.A., Dove W.F.;  
RT "Multiple intestinal neoplasia caused by a mutation in the murine  
RL homolog of the APC gene.";  
RN Science 256:668-670(1992).  
RP ERRATUM.  
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,  
RA Luongo C., Gould K.A., Dove W.F.;  
RL Science 256:1114-1114(1992).  
RN [3]  
RP SEQUENCE OF 1-45 FROM N.A.  
RC STRAIN=BALB/C; TISSUE=Liver;  
RA Dicker F., Lambert S., Reitmaier A., Ballhausen W.G.;  
RT "The murine APC gene: alternative splicing of 5' untranslated  
RN region segments.";  
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP ALTERNATIVE SPLICING.  
RX MEDLINE=94061824; PubMed=8242607;  
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;  
RT "APC gene messenger RNA: novel isoforms that lack exon 7.";  
RN Cancer Res. 53:5589-5591(1993).  
CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1  
CC and participates in Wnt signaling. APC activity is correlated with  
CC its phosphorylation state (by similarity).  
CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds  
CC axin (by similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=4;  
CC Name=1;  
CC IsoId=Q61315-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q61315-2; Sequence=VSP\_004116;  
CC Name=3;  
CC IsoId=Q61315-3; Sequence=VSP\_004117;  
CC Name=4;  
CC IsoId=Q61315-4; Sequence=VSP\_004116, VSP\_004117;  
CC -!- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,  
CC lung, brain, stomach, intestine, testis and ovary.  
CC -!- PTM: Phosphorylated by GSK3B (by similarity).  
CC -!- SIMILARITY: Contains 7 ARM repeats.  
CC  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M88127; AAB59632.1; -;  
DR EMBL; U02937; AAA03443.1; -;  
DR PIR; I49505; I49505.  
DR HSP; Q02248; 3BCT.  
DR MGD; MGI:88039; Apc.  
DR GO; GO:0005737; C:cytoplasm; IDA.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR GO; GO:0008013; P:beta-catenin binding; IDA.  
DR GO; GO:0009952; P:anterior/posterior pattern formation; IMP.  
DR GO; GO:0009798; P:axis specification; IMP.  
DR GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.  
DR GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.  
DR InterPro; IPR008938; ARM.  
DR InterPro; IPR000225; Armadillo.  
DR Pfam; PF00514; Armadillo\_seg. 4.  
DR SMART; SM00185; ARM; 5\_seg; 4.  
DR PROSITE; PS0176; ARM\_REPEAT; 1.  
KW Wnt signaling pathway; Anti-oncogene; Phosphorylation;  
KW Alternative splicing; Repeat; Coiled coil.  
FT DOMAIN 1 61 COILED COIL (POTENTIAL).  
FT DOMAIN 125 245 COILED COIL (POTENTIAL).  
FT DOMAIN 1 728 LEU-RICH.  
FT REPEAT 451 493 ARM 1.



FT REPEAT 503 545 ARM 2.  
FT REPEAT 546 589 ARM 3.  
FT REPEAT 590 636 ARM 4.  
FT REPEAT 637 681 ARM 5.  
FT REPEAT 682 723 ARM 6.  
FT REPEAT 724 765 ARM 7.  
FT DOMAIN 739 2834 SER-RICH.  
FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 1556 1871 ASP/GLU-RICH (ACIDIC).  
FT DOMAIN 1864 1891 HIGHLY CHARGED.  
FT VARSPLIC 243 276 Missing (in isoform 2 and isoform 4).  
FT VARSPLIC 310 410 /FtId=VSP\_004116.  
FT VARIANT 120 120 T -> A (IN STRAIN CAST/EI).  
FT VARIANT 493 493 Y -> I (IN STRAIN CAST/EI).  
FT VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).  
FT VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).  
FT VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).  
FT VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).  
FT VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).  
FT VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).  
FT VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).  
SQ SEQUENCE 2845 AA; 311086 MW; 145CAV3CF570A499 CRC64;  
Query Match 2.2%; Score 204; DB 1; Length 2845;  
Best Local Similarity 19.9%; Pred. No. 0.17;  
Matches 412; Conservative 272; Mismatches 808; Indels 580; Gaps 97;  
4 KSLGTEHKAHVTAHNPVGHVALQGGSSSSPONAAASLAABGKRGKMPRIHQPSTA 63  
242 RSSQSRDAASHAGRQHEGHAESNTAASSGQSPAT-----RYDHETAS 288  
64 ADGISAHQKKSFLRGCLGKTK--KFSRSPAGQPGTTHSKGATLRDLARDGQTOH 120  
289 VLSSGTHSAPR--RLTSHLGTKVMYSLMLG-----THDKDMGRTLLAM-- 335  
121 EAAAPDAARLTSGGVKRRNDDMAGRP--VKGSGGEDKVPTQKRHLNNFGMRQT 177  
336 -SSQSDSCISMRSQSGC-----LPLLIQLLHGNDKDSVLLGNSRG----- 373  
178 MLSKMAHPASAGADRIQHPPIPGSHHEIKKEPVGSGSKATTAHADRVIEAIEDDSE 237  
374 --SKEARARASAAALHNIHSQDPDKRRREIR-----VLHLEQIRAYCETCWE 420  
238 FOOLHQRLARENPPOPKLGVATPISARFQPKLTAVAESVLEGTDTTQSPKQSM 297  
421 WQEAHEGMDQDKMPAPVEHQIC----- 445  
298 KGSAGVTPPLAVTLDKGLQAPDNPALNTLLKQTLGKDTQHYLAHASSD--GSQHL- 354  
446 -----PAVCVLMK--LSFDEEH--RHAMNELGGLQIA 474  
355 -LLDNKGHLFDIKSTATSYSVLENSHPGEIKGLQAAGTGSVVDGSGKISLGSQTOSH 413  
475 ELLQVDCMEGL--TNDHYSVTLRYAG-----MALTNLTFFGDVAN 513  
414 NKTMLSQPGEAHRSLLGIWCHPAGAPQGESIRLHDDKIHLHPGLVWQSAKDQTHS 473  
514 KATLCSMKG--CMRALV-----AQLKSESEDLQOVIASVLIR--NLS-WRA---DVNS 557  
474 QLSRQADGLKLYAL-----KDNRTLQ-----DNKSKBLVDKIKSYSV-- 513  
558 KKTLEVGSKVAMECALEKVEKSTLSVLSALNLSAHCENKADICAVDGAFLVGT 617  
514 ----DQRGQVAILTDTGRHOMSTMPSLDASPESHISLSLHFADAHQCLL-HGKSELEAQ 568  
618 LTYRSQNTNLAIRES--GGGILRVSSLIATNEDHQRI-LRENNCLQTLQLHLK----- 668  
569 SVATSHGLVVADSEGLF--SAALPKQGDGNELKMKAMPOHALDEHGHQHSIGFFHD 626  
669 ----SHSLTIVSNACGTLWNLSARNPKD-----QEALED-WGAVSMLNLIHS 711

QY 627 DHQQL---NALVKNNFRQQHACPLGNDHQPHQGNWLTDLALVIONQLGLHHTNPPHEI-L 682  
Db 712 KHRMIAMGSAALRNLMANRPYKDKANIMSPGSSLP-----SLHVRKQKALEAEL 762  
QY 683 DMGHLGSLAQEGKLIHYFDQLTGKWTGAESDCQL--KKGL--DGAAYLLKDGVEKRLNIN 739  
Db 763 DAQHLSET-----FDNIDMLSPKASHRSQRHKQNLGYGVAFDANRHHDSRSNFN 813  
QY 740 QSTSYKHGTENVFSLPHVRNKPEPPDALQGLAKDDKAQAMAVIGNVKYLALTEKGDIRS 799  
Db 814 TGNMTVLSPLYNTVLPSSSSSRGSLDSSRS-EKDRSLERERGLGLSAYHPTTIENTACTSS 872  
QY 800 FQKPGTQQLERPAQTLSREGISGELKDIIHVDHKQNLIALTHEGEVPHQPREAMQNGAES 859  
Db 873 ---KRGIIQITTTAAQIAK---VMEEVSAIHTSQDDRSSASTTE---PHCVADDRSAARRS 923  
QY 860 SSWHKLK---LPQSESKLKLDM-----SHEHKPIATFED--GSQHLQKAGGWH 903  
Db 924 SASHTSNTVNTFKSENSNRKTCMPYAKVKRSSNDLSNSVTSSDGYGKRGQMKP-SVE 982  
QY 904 AYAAPERGPLAVGTSGSTVFNRL-----MQGVKGVIPGSGGLTVKLSAQ---TSGMTGA 955  
Db 983 SYSEDDSEKFCYGOYPADLAHKIHSANHMDNDGELDTPIYNSLYSDEQLNSGRQSPS 1042  
QY 956 EGRKVS-----KFSERIRAYAFN---PTWSTPRPIKNAAYATQHCWQREGLKPL 1003  
Db 1043 QNERWARPKHVIDEIKQNEQROARSQNTSYPVYSENTDDKHLKFPQHFQOQ--ECVSP- 1099  
QY 1004 YEMQGALIKQLD---AHNVYRNAPODLOSLETLDLGEHGAELNDKMRFRDELBQSA 1059  
Db 1100 YRSRGTSGETNRMGSHAINQNVNOSLQEDDYEDDKPTNYSERYSEEHQHEEBERPT 1159  
QY 1060 TRSVTVLQGHQV-----LKSNGEINSEFKSPGKALVQSFNVRNSGQD-----LSKS 1107  
Db 1160 NYSIKYNEEKHVDQPIDYSLKYATDILSSQKPS-----FSFKNSAQSKEHLSPS 1213  
QY 1108 LOQAVHATPSPASKIQSMLGHFVSGVDMSHQKG--EIPLGRQRDPNDKALTAKRLI 1164  
Db 1214 SENT--AVPPS-NAKRNQL-RPSSAQNRNGTQKGTTCVVP-----SINGETIQT--YCV 1262  
QY 1165 LDT-VTIGELHELADKAKLYSDHKPDADQIKOLRQOFTDLREKRYESNPVKHYTDMGFT 1223  
Db 1263 EDTPICFRSCSSLSLSS--ADDEIGCDQTTQEADSANTLQTAEVKENDVTRSAEDPATE 1320  
QY 1224 NKALEANYDAVKAFINA-----FKHEHGVNLTTRTVLESQGSAAELAKK-----LKNLT 1272  
Db 1321 VPAVSQARAKPRLQASGLSSSTRINKAVESSGAKSPSKGCAQTPKPPHVVQETP 1380  
QY 1273 L-----SLDSGSMFSRSY-----GGGVSTVFVPT----- 1298  
Db 1381 LVFSRCTSVSLDSFESRSIASSVQSEPCSMVSGIISPDLPSQQTMPPSRSKTPPP 1440  
QY 1299 -----LSKKVPVPIPGAGITLDRAYNLSFRSTG---GLNVSGFRDGGVSGNIMVAT 1348  
Db 1441 PPQTVQAKREVPKSKVPAAE-----KRESGPKQTAVNAVOR-----VQVLP 1482  
QY 1349 GHVMPYMTCKKTISAG-NASDWLSAKHKISP-----DLRIGAAVSGTLOQTQLNSLKFK 1401  
Db 1483 DVTLLHFATESPDGFCSSSLSALSLEDEFFQKQVELAIMPPVQ----- 1528  
QY 1402 LTEDLPGFTHGLTHGTLTPAELLQKGIHQMGKSKLTFS---VOTSANLDEL----- 1451  
Db 1529 --END-----NGNETESEQPEESNQDEVEKPDSEKOLLDDSDDDDEILEBECI 1577  
QY 1452 -----PAGINLNEGSKPNGVTVARVSAGLSASANLAAGSRERTTSGQFSGTTSAS 1502  
Db 1578 ISAMPTKSSRKAKLAQATASKLPPPVARKPSQLPVYKLLPAQNRLQAKHVSF----- 1630  
QY 1503 NNRPTFLNGVAGANLTAALGVAHSSTHEGKPGVGPFAFTSTNVSAALADNRSTSQISL 1562  
Db 1631 -----TPGDDVFRVYCVGETPI-----NPFSTATSLDLTIESPPN 1665  
QY 1563 ELKRAEPVTSN-----DISELTSTLKGKFKDSATTKMLAAL-KELDDAKPAEQLHILQ 1615

Db 1666 ELATGDSVAGIQSGEFERDITPTGRSTDDAQRGKISSIVTPDLDDNK-AREGDILAE 1724  
QY 1616 HFSKDVGVDEREYARNKLVIRQAADSHS-----MELGSASHSTT 1659  
Db 1725 CINSAMPKG-KSHKPRF-VKIMDQVQQAASSTSSGANKNQVDTKKKPTSPVKMPMPQTE 1782  
QY 1660 YNNLSRINNDGIV-----ELLKHFDPAALPA--SSAKRLGEMM--NNDPALKDII----- 1705  
Db 1783 YRTRVRKNTDSKVNVTETFSNDKDSKKPSLQTNAKAFNKLNNEDVRVGTGTFALDSPH 1842  
QY 1706 --KQLGSTFP-----SSASVSMELKDLRLRQTEKAILDKGVKREEVGVLEFQDRNNLR 1755  
Db 1843 HVTPIRGTPCYFSRNSLSDLDVDSREKAE--LRKGESKDS-----EAKVTCR 1895  
QY 1756 VKSVSVQSVSKSE--GFNTPALLLGTSNAAAMSMERNIGTINFKYGQD----- 1802  
Db 1896 PFNPSQQQAASQSASIKHPA---NRAQSKPVLOKQPTFPQSSKQDGPDRGAATDEKLNQL 1952  
QY 1803 --QNTFRFTLEGGIAQAANPQVASALTDLKKK 1832  
Db 1953 AIENTVPCFSRNSL-----SSLSDIDQE 1976

## RESULT 13

BPEA HUMAN  
ID BPEA HUMAN STANDARD; PRT; 5171 AA.  
AC Q94833; Q8N1T8; Q8WKK9; Q96AK9; Q96DQ5; Q9H555;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Bullous pemphigoid antigen 1, isoforms 6/9/10 (Trabeculin-beta)  
DE (Bullous pemphigoid antigen) (BPA) (Hemidesmosomal plaque protein)  
DE (Dystonia musculorum protein).  
GN BPAG1 OR DMH OR DT OR KIAA0728.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A. (ISOFORM 6), AND TISSUE SPECIFICITY.  
RC TISSUE=Keratinocytes;  
RX MEDLINE=21839111; PubMed=11751855;  
RA Okumura M., Yamakawa H., Ohara O., Owaribe K.;  
RT "Novel alternative splicings of BPAG1 (bullous pemphigoid antigen 1)  
RT including the domain structure closely related to MACP (microtubule  
RT actin cross-linking factor)";  
RL J. Biol. Chem. 277:6682-6687(2002).  
RN [2]  
RP SEQUENCE OF 1342-5171 FROM N.A. (ISOFORM 10).  
RC TISSUE=Ductenium;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altshul S.F., Zeeberg K.H., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T.J., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Rulyk S.W.,  
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]

RP SEQUENCE OF 1702-4156 FROM N.A. (ISOFORM 6).  
RC TISSUE=Brain, Placenta, and Tongue;  
RX Ninomiya K., Wagatsuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,  
RA Furuuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,  
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Murakawa K.,  
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,  
RA Suzuki Y., Sugano S., Senoh A., Mizuguchi H., Inagaki H.,  
RA Nakagawa S., Sato H., Kikuchi H., Masuho Y.,  
RA Nagai K.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 3535-5171 FROM N.A. (ISOFORM 6).  
RC TISSUE=Brain;  
RX MEDLINE=22158633; PubMed=12168954;  
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;  
RT "Construction of expression-ready cDNA clones for KIAA genes: manual  
RT curation of 330 KIAA cDNA clones.";  
RL DNA Res. 9:99-106(2002).  
RN [5]  
RP SEQUENCE OF 4107-5171 FROM N.A. (ISOFORM 6).  
RC TISSUE=Brain;  
RX MEDLINE=99087487; PubMed=9872452;  
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XI.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 5:277-286(1998).  
RN [6]  
RP SEQUENCE OF 4031-5171 FROM N.A. (ISOFORM 9).  
RC TISSUE=Keratinocytes;  
RX Smith M.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 4031-5171 FROM N.A. (ISOFORM 9).  
RC TISSUE=Keratinocytes;  
RX MEDLINE=92011493; PubMed=1717441;  
RA Sawamura D., Li K., Chu M.-L., Uitto J.;  
RT "Human bullous pemphigoid antigen (BPAG1). Amino acid sequences  
RT deduced from cloned cDNAs predict biologically important peptide  
RT segments and protein domains.";  
RL J. Biol. Chem. 266:17784-17790(1991).  
CC -!- FUNCTION: Cytoskeletal linker protein. Anchors keratin-containing  
CC intermediate filaments to the inner plaque of hemidesmosomes. The  
CC proteins may self-aggregate to form filaments or a two-dimensional  
CC mesh (By similarity).  
CC -!- SUBUNIT: Homodimer (By similarity).  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=10;  
CC Name=6; Synonyms=EA;  
CC IsoId=Q94833-2; Sequence=Displayed;  
CC Name=1;  
CC IsoId=Q03001-1; Sequence=External;  
CC Name=2;  
CC IsoId=Q03001-2; Sequence=External;  
CC Name=3; Synonyms=1e;  
CC IsoId=Q03001-3; Sequence=External;  
CC Name=4;  
CC IsoId=Q03001-4; Sequence=External;  
CC Name=5;  
CC IsoId=Q03001-5; Sequence=External;  
CC Name=7; Synonyms=EB;  
CC IsoId=Q8WKK8-2; Sequence=External;  
CC Name=8;  
CC IsoId=Q03001-6; Sequence=External;  
CC Name=9;  
CC IsoId=Q94833-3; Sequence=VSP\_005068, VSP\_005069;  
CC Name=10;  
CC IsoId=Q94833-1; Sequence=VSP\_005066, VSP\_005067;  
CC Note=No experimental confirmation available;  
CC -!- TISSUE SPECIFICITY: Highly expressed in skeletal muscle and

CC cultured keratinocytes.  
 CC -!- DISEASE: BPAG1 is an autoantigen of bullous pemphigoid  
 CC [MIM:600088], an autoimmune subepithelial skin blistering disease.  
 CC -!- SIMILARITY: Belongs to the plakins or cytokeratin family.  
 CC -!- SIMILARITY: Contains 2 EF-hand calcium-binding domains.  
 CC -!- SIMILARITY: Contains 1 SH3 domain.  
 CC -!- SIMILARITY: Contains 29 spectrin repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcements/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF400226; AAL62061.1; -;  
 CC EMBL; BC016991; AAL6991.1; -;  
 CC EMBL; AK055189; BAB070870.1; ALT INIT.  
 CC EMBL; AK094883; BAC04449.1; ALT INIT.  
 CC EMBL; AK096713; BAC04848.1; ALT INIT.  
 CC EMBL; AB018271; BAA34448.2; -;  
 CC EMBL; ALJ37008; CAC12899.1; -;  
 CC HSPF; P02631; IRRO.  
 CC Genew; HGNC:1090; BPAG1.  
 CC MIM; 113810; -;  
 CC MIM; 600088; -;  
 CC GO; GO:0005737; C:cytoplasm; ISS.  
 CC GO; GO:0005200; F:structural constituent of cytoskeleton; ISS.  
 CC GO; GO:0045104; P:intermediate filament cytoskeleton organiza. .; ISS.  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR003108; GAS2.  
 CC InterPro; IPR001452; SH3.  
 CC InterPro; IPR02017; Spectrin.  
 CC Pfam; PF000036; ehand; 2.  
 CC Pfam; PF02187; GAS2; 1.  
 CC Pfam; PF000018; SH3; 1.  
 CC Pfam; PF00435; spectrin; 29.  
 CC ProDom; PD000012; EF-hand; 1.  
 CC SMART; SM00054; EFh; 2.  
 CC SMART; SM00243; GAS2; 1.  
 CC PROSITE; PS00018; EF\_HAND; 2.  
 CC Antigen; Actin-binding; Coiled coil; Repeat; SH3 domain;  
 CC Structural protein; Cytoskeleton; Cell adhesion; Calcium;  
 CC Calcium-binding; Alternative splicing.  
 CC REPEAT 264 341 SPECTRIN 1.  
 CC REPEAT 349 444 SPECTRIN 2.  
 CC REPEAT 450 550 SPECTRIN 3.  
 CC FT DOMAIN 564 616 SH3.  
 CC REPEAT 935 1031 SPECTRIN 4.  
 CC REPEAT 1481 1581 SPECTRIN 5.  
 CC REPEAT 1715 1814 SPECTRIN 6.  
 CC REPEAT 1930 2008 SPECTRIN 7.  
 CC REPEAT 2071 2174 SPECTRIN 8.  
 CC REPEAT 2184 2282 SPECTRIN 9.  
 CC REPEAT 2294 2382 SPECTRIN 10.  
 CC REPEAT 2404 2502 SPECTRIN 11.  
 CC REPEAT 2513 2611 SPECTRIN 12.  
 CC REPEAT 2839 2940 SPECTRIN 13.  
 CC REPEAT 2950 3047 SPECTRIN 14.  
 CC REPEAT 3059 3156 SPECTRIN 15.  
 CC REPEAT 3168 3268 SPECTRIN 16.  
 CC REPEAT 3277 3376 SPECTRIN 17.  
 CC REPEAT 3386 3484 SPECTRIN 18.  
 CC REPEAT 3495 3594 SPECTRIN 19.  
 CC REPEAT 3716 3816 SPECTRIN 20.  
 CC REPEAT 3825 3925 SPECTRIN 21.  
 CC REPEAT 3935 4035 SPECTRIN 22.  
 CC REPEAT 4044 4142 SPECTRIN 23.  
 CC REPEAT 4153 4250 SPECTRIN 24.  
 CC REPEAT 4262 4362 SPECTRIN 25.  
 CC REPEAT 4372 4470 SPECTRIN 26.  
 CC REPEAT 4480 4578 SPECTRIN 27.

FT REPEAT	4596	4684	SPECTRIN 28.
FT REPEAT	4732	4790	SPECTRIN 29.
FT CA BIND	4798	4810	EF-HAND 1 (POTENTIAL).
FT CA BIND	4834	4846	EF-HAND 2 (POTENTIAL).
FT DOMAIN	126	160	COILED COIL (POTENTIAL).
FT DOMAIN	312	377	COILED COIL (POTENTIAL).
FT DOMAIN	402	447	COILED COIL (POTENTIAL).

Query Match 2.1%; Score 202; DB 1; Length 5171;  
 Best Local Similarity 18.1%; Pred. No. 0.53;  
 Matches 364; Conservative 310; Mismatches 720; Indels 616; Gaps 94;

Qy	143	DMAGRPVMYKGSSEKVPYTOQ---KRHLNFMGQMTMLSKMAHPASANAGDLQHSPP	199
Db	1138	EKAGKPPF---SKQKISSEISTKKEQLSEALQTIQLFLAKHGDKMTDEERNELEKQVK	1193
Qy	200	HIPGSHHEIKEEPPVGSTSKATTAHADRAVEITAEQDDSEFOQLHQOR-----LAREEN-	252
Db	1194	TLOESYNLFSSESLKQLQESQTSQGDVKE---EKIVAEQOEYKBEKLOGICDILLTQENR	1250
Qy	253	---PPQPPKLGVAIPISARFPK-----LTVAESVLEGTDTTQSPKPSQMLKSGSA	302
Db	1251	LIGHQEAFMIGDGTVELKYYQSEELQDKMQSAQALAEVVKNTENFLK-ENGEKLSQE	1309
Qy	303	GVTPPLAVTLDKGLQL-----APDNPALNTLLKOTLGKDTQHYLAHASSDGSQHL--	354
Db	1310	DKALIEQKNEAKIKCEQLNKAQSKKELDKVVTVAIKEETKVAQVQLEBSKTKIEN	1369
Qy	355	LLDNKGHLFDIKSTATSYSVLHNSHPGFIKGLAQAGTGSVSDGSKGKSLGSG-----	408
Db	1370	LLD---WLSNVDKDSERAGTKH-----KQVTEQNGTHFQEGDGKS---AIGSEDEWNG	1416
Qy	409	-----GTQSHNKTMLSQPGA-HRSLTLGIWHPAGAARPOGESIRLHDDHIHI	456
Db	1417	NLLETVDVGQVGTQENLNQYQVKAQHEKITS---OHOAVIITASAQVLLKQGOYL	1473
Qy	457	-----LHPGLVMQSAKDQ--THS-----OLSRQ--	478
Db	1474	SPEKEKQLQNMKELKVHYETALAESEKMKMLTHSLQEELEKFPDADVTEPEHMLQSEQE	1533
Qy	479	-----ADGKLYALKONRTL-QNLSNK-----SSEKLVDKIKSYSDVQGOV	519
Db	1534	LENLEAGAGDINGIMTKLKRQKSFSEVDVISHKGLDRYITISGNRVLEAAKSCSRDGGKV	1593
Qy	520	ALLTDTGPRHKMSIMPSLDASPESHISL-----SLHFADAHQGLLHGK	562
Db	1594	----DTSATHR-EVQRKLDHATDFRSLSYKCNVGNLXDLVDKYQHYEDASGGLAGL	1648
Qy	563	SELEA-----OSVAISHGRVVDSEKGLFSAAIKQGDGNEKMKAMPQHALDBHFG	615
Db	1649	QACEATASKHLSPIAVDPKNLQRLQLEETKALQGISSQVAVVE-KLKKTAEVLDA---	1704
Qy	616	HDHQISGFFHDDHCOLNALVKNPNFRQOHACPLGNDHQHFHGWNLTDALVDNQLG---	671
Db	1705	-----RGSLL-----LPAXNDIQ-----KTLDIDVGRYEDL	1729
Qy	672	HHTNPEPEILDMHGLSLALQEGKLHYFDLTQKGTG-AESDCK-OLKGLDGAAYLLK	729
Db	1730	SKSVNERNEKQITLFRSLSVQDGLDEMLD-----WNGVSESLKEQGVPLNSTA--LQ	1782
Qy	730	DGEVKRLINQOS-----TSSIKHGTENVFSLPHVRNKPFGDALQGLNKDKAQAMAVIGV	785
Db	1783	DIISKIMLEQDIAGROSSINAMNEKVKFMTETD-PSTASSLQAKMKDLARPSEASHK	1841
Qy	786	NK-YLALTE--KGDIRSFOIKPGTQO--LERPAQTLRSREGISGELKDI-----	828
Db	1842	HKETLAKMEELKTKVELFENLSKLQTLFTLTKTQALTEVDVPG--KDVTELSTQYMQBSTS	1899
Qy	829	-HYDHNKONLYALTHEGRVPHQ-PREAWQNGAESSSWHKLALPOSES---KLKSL-DMSHE	882
Db	1900	EFIEHKKHL-----EVLHSLIKEISSHGLPSDK--ALVLEKINNLSKFKEMEDTIKE	1950
Qy	883	HKTIATFEDSGSHQLKAGGWHYAAAPERGLAVGTSGTSQTVFNRLMQGVK-----KVI	936

Db 1951 KKEAVT---SCQQLDA-----FQVLVKSLSWIKETTKV 1983  
QY 937 PGSGTLVKLSAQGTGAGTKVSKFSPR-----IRAFNFTWSTPRPI-- 983  
Db 1984 P--IVQSPFGAEDLGSLEDTKLQKWSUKTPEIOKVNNSGISLNLISAVTTPAKAIA 2041  
QY 984 -----KNAAYATQGMQREGKPL-----YMQGALIKQLDAHVREN 1022  
Db 2042 AVKSGGAVLNGEGTATNTEFW-ANKGLTSIKDMTDSHGVEDLGLLKDKIA----- 2094  
QY 1023 APQDLOSLETLDLGEHGAELNDMKRPRDELEQSAIR-SVTVLQHQGVLSKNGEINS 1081  
Db 2095 -----ELNLTLSKL---QKAQESSAMQWLQKMNKTATKWQTPAPTDEAVKTVQEQNK 2147  
QY 1082 EFKPSFGKLVOSFNYSQDLSKLSQAQVHATPPSAES-KLSQMLGHFVSAGVDMSHQ 1140  
Db 2148 SFEEA-----LQK--NVNKV-QLKDKLTLEENPDTPAPRWQML-----TEIDSKWQ 2195  
QY 1141 K-GEIPLGRQ-----DPNDKT-----ALTKSRLIILDTVTIGELHELADKAKIVSHKPD 1189  
Db 2196 ELNQLTIDRQKLEESSNNLTQFTVEAQLKQWLVEKELMVSVLGPLSIDENMLNQTQ-- 2253  
QY 1190 ADQIKLROQFDTLREKRYE-----SNPVKHYTDMGFTHK--ALEANYDAVKAF 1237  
Db 2254 -QQVQILLQEFAT-RKPYEQEULTAAGQGLSRPGEDPSLRGIVKEQLAAVTKWDSLTGQ 2311  
QY 1238 INAFKKEHGVNLTTRTVLESQGSALAKKLKNTLLSLDSGESMSFSRSGVGGVSTVFPV 1297  
Db 2312 LS-----DRCWIDQAIKVSQYQSLSLSLDKLSLDLN----- 2345  
QY 1298 TLSKVPVPVPIGAGITLDRAYNLSFRTSGGLNVSGFRDGGVSGNIMVATGHVMPYMT 1357  
Db 2346 KLSSSLAVSTHEDA-----MNQLETAQMKQEIQ 2375  
QY 1358 GKTSAGNASDWLSAXHKLSPDLRIGAAVSGITLQTLQNSLKPKLTDELPFIHGLTHG 1417  
Db 2376 QEKQKIKVAQALCEDLSALVKEYLKAELSRQLEGILKS---FKDVEQAENHVQHLQSA 2432  
QY 1418 TLTTPABLLQKLEHOMQKQSKLTFSDVTSANLDLAGINLEDGSPKNGVTARVSAGLSA 1477  
Db 2433 CAS-----SHQFOQMSR-----DFQAWLDTK-----KEQNKSHPIAK-----LDV 2469  
QY 1478 SANLAAGSRERTTSQFGSTTSASNRRPTFLNGVGAGANLTAALGAHVSSTHEGKPVGI 1537  
Db 2470 LESLIDKHDK-----FSKTLTAQSH--MYEKRTIAEGENLLK----- 2504  
QY 1538 PFAFTSTNVSAAAL-----DNRTSQSISLELKRAEPTVS----- 1572  
Db 2505 ---TQGSKAALQQLNTIKTNWDTFNKQVKERENKJESLEKALKYKEQVETLWPWID 2560  
QY 1573 ---NDISELTSTLGRHKFDKDSATTKMLAALKELDDAKPAQLHLOQHFSA----- 1619  
Db 2561 KCONLEEKFKCLDPAGENSIAKLSLQKEMD-----QHFGVLLNNTANS 2608  
QY 1620 -----KDVGVDEREYAVRNL-----KKLVI 1639  
Db 2609 LLSVCEIDKGVVTDENKSLIKQVDMVTEQLHKKFKCLENNTKQFKEFQEVSKESKRLQC 2668  
QY 1640 RQQAADSHSMELGSASHSTYNNLSRINNDGIVELHKKHFDALPASSAKRLGEMM----- 1695  
Db 2669 AKEQLDIHD-SLGSQAYSNNKLTMLQTKQSLQALKHQ-----VDLAKRLQAQDLVVEA 2720  
QY 1696 NNDPALKDIKIQLOQSPFSSASVMEL-----KDLREQ-TEKAILDG 1737  
Db 2721 SDSKGTSDVLLQVETIAQHESTLSQVDEKCSFLETQLOGIGHFRNTIREMFSQFAFDD 2780  
QY 1738 KVGREVGVLFPQDNNLRVKSVSQSKSEGF-----NTPALLIGTNSA--AMS 1787  
Db 2781 EL--DSMAPVGRDAETLQKQKETIKAFKLKLEALMASNDNANKTKOMLATERTSPDLVG 2838  
QY 1788 MERNIGTINFKYQO--DONTPRFTLEGGI 1815

Db 2839 IKRDLEALSQCNKLLDRAQAAREEQVEGTI 2868  
RESULT 14  
FHAB BORPE  
ID FHAB BORPE STANDARD; PRT; 3590 AA.  
AC P12255;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Filamentous hemagglutinin.  
GN FHAB OR BP1879.  
OS Bordetella pertussis.  
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
OC Alcaligenaceae; Bordetella.  
OX NCBI\_TaxID=520;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RX MEDLINE=90355839; PubMed=2388559;  
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
RT "Genetic characterization of Bordetella pertussis filamentous  
RT haemagglutinin: a protein processed from an unusually large  
RT precursor".  
RL Mol. Microbiol. 4:787-800 (1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;  
RX MEDLINE=22827954; PubMed=12910271;  
RA Parthill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,  
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
RA Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,  
RA Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,  
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,  
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,  
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;  
RT "Comparative analysis of the genome sequences of Bordetella pertussis,  
RT Bordetella parapertussis and Bordetella bronchiseptica".  
RL Nat. Genet. 35:32-40 (2003).  
RN [3]  
RP SEQUENCE OF 1-3261 FROM N.A.  
RX MEDLINE=89202384; PubMed=2539596;  
RA Relman D.A., Domenighini M., Tuomanen E., Rappuoli R., Falkow S.;  
RT "Filamentous hemagglutinin of Bordetella pertussis: nucleotide  
RT sequence and crucial role in adherence".  
RL Proc. Natl. Acad. Sci. U.S.A. 86:2637-2641 (1989).  
CC -!- FUNCTION: Evidence for a role in host-cell binding and infection.  
CC -!- SUBCELLULAR LOCATION: Surface.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M60351; AAA22974.1; --  
CC EMBL; EX640416; CAE42162.1; --  
CC PIR; S21010; S21010.  
CC InterPro; IPR008619; Fil haemagg.  
CC InterPro; IPR008638; Haemagg act.  
CC Pfam; PF05594; Fil haemagg; 18.  
CC Pfam; PF05860; Haemagg\_act; 1.  
CC Antigen; Hemagglutinin; Complete proteome.  
CC CONFLICT 507 508 KQ -> NE (IN REF. 1 AND 3).  
CC CONFLICT 1454 1454 A -> P (IN REF. 1 AND 3).  
CC CONFLICT 3574 3590 VEDIGKNRYVFYETNK -> SRISAARTTGSSMKPTNR  
CC (IN REF. 1).  
CC SEQUENCE 3590 AA; 367519 MW; C00BD8E22C9DB41D CRC64;

Query Match 2.1%; Score 199.5; DB 1; Length 3590;  
 Best Local Similarity 19.3%; Pred. No. 0.39;  
 Matches 426; Conservative 268; Mismatches 793; Indels 725; Gaps 114;

Qy	5	SLGTEHKAHVTAHNPVGHVALQOGSSSPQWAAASLAAEKNRKMPIRIHOPSTAA	64
Db	1379	SLDIKKGGAQVTVAGRYAEHGEVSTQGDYTVSAD--AJALAAQVTVRG--GA	1426
Qy	65	DGISAHQKQKSFSLRGCLGKTKFSRSAPQGPQGTTH--SKGATLLDILLARDGTOHEAA	123
Db	1427	ANLISRHDTRESNKR-LMGLPQVNVAGGAVNTGNLKVREGVTV--TAASFNETGAEMV	1483
Qy	124	APDAARLTRSGVKRRNMDMAGRPMVK-----GGSG	155
Db	1484	A-KSATLTTSGAARN-----AGKMQVKEAATIIVAASVSNPGTFTAGKDIITVSRGGFDN	1536
Qy	156	EDKVPTQ-----KRHLNN-----EQMRQTMLSKMAHPASANAGDRLQHS	197
Db	1537	EKMESNKDVIKTEQFNGRVLDAKHDLTVTASQADNRSLKAGHDFTVQA--QRIDNS	1595
Qy	198	PPHIPGSHHEIKBPVSGTSKATTAAHADRVEIAQEDDDSEFQQLHQOQLARERENPPQPP	257
Db	1596	GTMAAGHDATLKAHLRNTGVVACH-----DIHIINSKLENTGRVDA	1639
Qy	258	KLGVATPISARFQPKLTAVAE-----SVLEGT-----TTQSPLKP--QSMILKSGAGVTP	306
Db	1640	RNDIALDV-ADFTNTGSLYASHDATLTLAQGTQRDVLVVDQPHILPVAEGTTLRVKAKSLTT	1698
Qy	307	LAVTLDKGLK---OLAPDNPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLLDNKGHL	362
Db	1699	EIETGNPGLTAEVOENTDNQAI-----VVGKDLTLSSA-HGNVANEANALLWAAGEL	1751
Qy	363	FDIKSTATSYSLVHNSHPGETKGKLAQAG-----TGSVSVGKSGKISLG-----SGTQS	412
Db	1752	-----TVKAQNI TNK-----RAALIEAGGNARLTAVALINKLRIPAGEDMHLDPRI	1800
Qy	413	HNKTMLSQPGBAHS-----LLTGIOHPAGAARP-----	442
Db	1801	ENTAKLS--GEVQRKGVDQGGGEGHRWSGTGYVNYMLRAGNKGKAGTIAAPWYGGDLTA	1858
Qy	443	-----QGESIRLH-----DUKIHLHPGLVQWQSAKDTHSLSRQADGK-----	482
Db	1859	EQSLIEVKOLYNAGARKDEHRLHNE--GVIQAGG---HGHI GGDVDRNSVVRTVSAM	1913
Qy	483	-----LYALKDNRTLQNLSDN-KSSKSLVDKIKSYSDVQGVQVAILTDTTPGRHKM	531
Db	1914	EYFKTLPVSLTALDNRAGLSPAATWQSTYELLD-----YLLDQNRVEYIWLGYTYTEW	1969
Qy	532	SI-----MPSLDASPESHISLSLHFADAHQ-----LLHG---KS	563
Db	1970	SVNTLKNLDLGYQAKPAPTAPMPKAPK--LDLRGHTLESAGRKIFGEYKKLQGEYEKA	2027
Qy	564	ELEAQSV-AISHGLVVDVADSGKLFSAAI-----PKQGDGNELKMKAMPQHALDEHFGH	616
Db	2028	KMAQVAEYGEATRRVHDQGLQRYGKALGMDAETKEVDG-----IIQEFAA	2075
Qy	617	DHOISGFFHDDHGQLNA---LVKNNFRQ--HACPLGNDHPHGPWNLTDLALVIDNQLG--	670
Db	2076	DLRTVYAKAQDAQIDAETDKVAQYKSIDAVRL---QAQPG-RVTLAKALSAAIGAD	2131
Qy	671	---LHHTN-----PEPHEILDHMGHLSLALQEGKHYFDOLTKGM	707
Db	2132	WRALGHSQLMQRMQDFKAGKRGAEIATFYKPEQTVLAAG--AGLTLSNGAIH-----NGE	2183
Qy	708	TGAESDCKQ---LKKG-----LDGAAYLLKD-GEVKRLINQSTSSIKHGTENTVPSLPHV	758
Db	2184	NAAQNRGRPEGLKTAHSATSVSGSFDAIRDVGLEKRLDIDALAA-----VLVNPFI	2236
Qy	759	-----RNKPE-----PGDALQGLNKDDKAQAMAVI	783
Db	2237	FTRIGAACATSLADGAGPALARQAPETQGMVDARGLSADALASLASLADAAQGLEVS	2296
Qy	784	GVNKYLALTEKGDIRSFOIKPFTQOLERP---AQTLSREGISGELKD---IHVDHKQNLV	837

Db	2297	G-RRNAQVADAG-----LAGPSAAPAAPVGAADVGVEPTGQVQDVVAVGLEQVPA	2348
Qy	838	ALTHEGEVHFOPREAWQNGAESHHKIALPQSSSKLSLSDMSHEHKFIATFEDSQSHQL	897
Db	2349	TVRVAPPAVALPR-----PLFETRIKFIQDQSKFY-----GSRYPF	2383
Qy	898	KAGQWHAAYAAPERGLAVGTSGSQTVE-----NRLMOQVKGV-IPGSGLTVKLSAQTTGG	951
Db	2384	EQIGY-----KPDRAARVAGDNYFTTLVREQVRALGGYESRLPVRGVALVAKLMDSAGT	2439
Qy	952	MTGAEGRKV-----SSKFSERIRAYAFNPTMTPTPIKNAAYATQHWGREGKLPKY	1004
Db	2440	VGKALGLKGVAPTAAQQLKQADRDFVWYVDVVIDQKVLAPRLYLTE---ATROQITDQY	2496
Qy	1005	EMOGALIK-----QLDAHNVRHNAQPDPLOSKLETLDLGHEGAELLNDMKRFRDELEQ	1057
Db	2497	AGGALLIAGSGDVTNTDGHV--SSVNGLIQGRSVKVDAGK-GKVYVADSKGAGGGIE-	2552
Qy	1058	SATRSVTVLGQHQGVLSNGEINSEFPKPGKALVQSPNVNRSGDLSKSLQQAHAHTPP	1117
Db	2553	-ADDEVDSGRDIGI--EGGKLR-----GKQVPLKADTVKVATSM	2589
Qy	1118	SAESKLOSLMUGHFVSAGVD-MSHQGEIPLQRQDPNDKTALT--KSLILDTVTVIGELH	1174
Db	2590	RYDDK-----GRLAARGDGDALDAQGQLHIEAKRLETAGATLKGKGVKLDVDDVKLG	2644
Qy	1175	ELADKAKLVSDHKPDADQIKOLRQOFDTLREKRVESNPVHYTDMGFTH--NKALEANY	1231
Db	2645	EAGSS-----YEN---KSTPLGSLFAILSSTETFNQ	2673
Qy	1232	DAVKAFINAFKKEHGVNLTTRTV-----LESQGSAAELAKKLNTLILSLDGSMSFSR	1285
Db	2674	SA-----HANHYGTIEIAGTLEGKMNLEIEGGSVDA---AHTDLSVARDARFKAA	2722
Qy	1286	SYGGGVSTVFPVTLTKKVPVPIPGAGITLDRAYNLSPRTS--GGLNVSVFGRDGVSN	1344
Db	2723	DFAHAHEKDVQRLSLGAKV---GAG---GYEAGFSLGSESGLEHAHAGR--GMTAGA	2771
Qy	1345	MVATGHDVMPYMT-----GKTSAGNASDMLSAKHKISPDL---RIGAAV	1386
Db	2772	EKVGYRASHEQSSSETEKSYRNANLNFGGGVSVEAGNVLDIGGA-----DINRNYGAA	2825
Qy	1387	SGTLQGTQLQNSLKFTELPEFPIHGLT-----HGTLPAPALLQKGI	1429
Db	2826	KGN-AGT-EAALRWAKKVESTKVVSEQTSSQSGSVSEVASTASARSLTLTAATLGDV	2883
Qy	1430	EHQMKQGSKL-----TFSVDTSANLDIRAGINLNBEDGSKPNQVTA-----	1469
Db	2884	AQNVEDGREIRGELMAAQVAEAATQLVTADTAA--VALSAGISADFSHRSRSTQNTQYL	2942
Qy	1470	-----RVSAGLSASANLAAGR--ERSITTSQGFSTTSASN	1503
Db	2943	GGNLSIEATEGDTALVQAKFGGGQVSLKKAASVNLMAAESTTFESYSESHNFHASDANL	3002
Qy	1504	NRPTFLNGVAGANLTAALGVVAHSSTHE-GKPVGIFPAFTSTNVSA--LALDNRTSQSIS	1561
Db	3003	GANAVQAGVGLG--LTAGMGTSHQITNETGK-----TYAGTSVDAANVSID--AGKDLN	3052
Qy	1562	LELKRAEPTVNSDISELTSTLGHKFKDSATTKMLAALKELDDAKPQBLHLQOHFSKAD	1621
Db	3053	LSGSRVR-----GKHVLDVEGDINATSKQ-----	3077
Qy	1622	VUGDER-YE-----AVRNKLKLVIRQQAADSHSMELGSA--SHSTTTNNLSRI	1666
Db	3078	---DERNYSNGGWDASAGVAIQN-RTLV-----APVGSAGFNFTNTEHDN--SRL	3122
Qy	1667	NNDGIVELL-----HKHFDAAALPASSAKRGLGEMMNDPALKDI IKQLQSTPSSASVS	1719
Db	3123	TNDGAAGVVASDGLTGHVKGDNALTGATIALDSKGN---LK-----VDGAVNA	3168
Qy	1720	MELKQGLREQTEKAILDKGVGREEVGVLFDQRNNLRVKSVSVSVSKSEGFNTPALLLG	1779

Db 3169 QNLKD-YRDK-----DGGSGGLNVG-----ISSTTLAPTGVAFG-----RVAG 3206

QY 1780 TNSRAMSERNIGTINFKYQDQNTPRBFTLEGGIAQANQVASALTDLKK 1831

Db 3207 EDYQAEQRATIDVG-----QTKDPARLQVGGVKGTLNQDAAQATVVQR 3250

## RESULT 15

CING\_MOUSE STANDARD; PRT; 1191 AA.

AC P59242;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Cingulin.

GN CGN.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahney J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield V.S., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,

RT human and mouse cDNA sequences."

RT generation and initial analysis of more than 15,000 full-length

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- FUNCTION: Probably plays a role in the formation and regulation of

the tight junction (nJ) paracellular permeability barrier (By

similarity).

CC -1- SUBUNIT: Homodimer. Interacts with TJP1/ZO-1 (By similarity).

CC -1- DOMAIN: Deletion of the ZO-1 interaction motif (ZIM) decreases but

does not abolish colocalization with ZO-1 (By similarity).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; BC042459; AAH42459.1; ALT\_INIT.

DR MGD; MGI:1927237; Cgn.

DR GO; GO:0005923; C:tight junction; NAS.

DR GO; GO:0003779; F:actin binding; ISS.

DR InterPro; IPR002928; Myosin tail.

DR Pfam; PF01576; Myosin tail\_1.

DR Tight junction; Coiled coil.

KW DONAIN 1 345 HEAD.

FT DONAIN 343 1150 COILED COIL (POTENTIAL).

FT DONAIN 1149 1191 TAIL.

FT DONAIN 357 830 GLU-RICH.

FT SITE 42 56 ZIM.

SQ SEQUENCE 1191 AA; 136446 MW; 43EA20F49AF5E7D9 CRC64;

Query Match

2.1%; Score 199; DB 1; Length 1191;

Best Local Similarity 19.4%; Pred. No. 0.073;  
Matches 265; Conservative 190; Mismatches 491; Indels 422; Gaps 66;

QY 16 TAAHNPVGH--GVALQGSSSS-----SPQNAAS--LAAEKNRGMKPRIH-QSTAA 64

Db 113 TSDPEPKDLHGLHRSQSASLTGLAFMSPNRSTLSLELAPKPTSSINTIDAPLSSV 172

QY 65 DGISAAHQOKSFSLRGCLGTKKFSRSPAQGGPGTTHSKGATLRDLARDGGTQHEAAA 124

Db 173 DSLINKFDSQKGGVGRGTGR--TITLPHQ-----RKRSQSLDSRLPRD---TREERH 223

QY 125 PDAARLTRSGGVKRRNMDMAGRPVMVGGSGEDKVTQOKRHQLNNFGQRMQTM----- 178

Db 224 QSANHWTR--GTKYDNHVDSSKNP-----SQKSPFSFSRSRQTDVWLQS 268

QY 179 LSKMAHPASAN-----AGDRLQHSPPHPIGSHHEIKKEPVGSGTSKATTAHADRVETIAQ 231

Db 269 FEETROPAMVQFKSTPDLLRDQRETAP--PGADHVKATYGI-----LRE 312

QY 232 EDDDSFQOLHQOQLARERENPPQPKLGVATPISARF---QPKLTAVAESVLEGTDTTQ 288

Db 313 GSSESEASVRRKVSLLVLEQMOP-----LGMVSPASTKALAGQAELTKMEELQKLD--- 364

QY 289 SPLKQSMMLKSGAGVT-----PLAVTLD--KGKLQAPNPPALNTLLKQTLG 335

Db 365 EEVKRQKLEPSRVGLERQLEKAEECHRLQELLERRKGEVQSSKELQNMKLLQOBEG 424

QY 336 -----KDOHYLAHASSD-GSOHL---LLDNK---GHLFDIKSTATSYSLHNSH 379

Db 425 LRHGLBAQVKELQKHKQSPDSGKESLKDLDLTRELLLELEGKQVVEQLRERE 484

QY 380 PGEIKGLKAQAGTG-----SVSDGKSGKISLGSQTQSH-----NKTMLSGPGE 423

Db 485 LTALKALKAEVASHQVEHVLQYQRTDTEQLRRSQMDATODHAALAEAKRMSSLVRE 544

QY 424 AHRSL-----LTGIWQ-----HPGAARPPQESIRLH-----DDKIHLHLPGLV 463

Db 545 LQRELEBETSEBTHGWSQMPKQKKEELRATKQELLQRLMEKEEBELGEMVELQRLDLEQ 604

QY 464 WQADADTHSOLSRQADGKLYALKDNRTLQNLSDNKSSEKLVKDKIKSYSDVQGGVAILT 523

Db 605 ARASTRDTH-----QVEELK-----KELRRTQELKELQAEQNOQ----- 639

QY 524 DTPGRHKMSIMPSLDASPESHISLSHFADAHQGLHKGSELBAQSVASHG-RLVVADS 582

Db 640 EVTGRHNRQVL-----EKQLAALREAD-----RGR-ELEQNLQQLKTLQQLRQDC 685

QY 583 EGKLFSAATPKQDGNELKMKAMPQALDEHFGHDHQISGFFHDDHCOLNALVKNNFRQQ 642

Db 686 E-----EASKAKVASETEAMML-----GQRATVETTLRET 716

QY 643 HACPLGNDHGFPGWNLTDALVIDNOLGLHHTNPPEHEILDMGLGSLAQEGKLHYFDQ 702

Db 717 QE-----ENDEFR-----RRILGLEQQLKEARGLAEGGEAVEARLSD-KVHRL-- 758

QY 703 LTKGWTGAESDCQKLLKGLDGAAYLLKDGVEKRLNINQSTSSIKHGTFENVPSLPHVRNKP 762

Db 759 -----EVEKQQLBEALNAAOE--EENGL-----AAAKRALE----- 787

QY 763 EPGDALQGLNKDDKAQAMAVIGNVKYL---ALTEKGDIRSFOIKPGTQQLERPAQTLSR- 818

Db 788 -----VRLDEAQRLARLQGOEQOQALNPALEBEGKQRE-ALRRSKAELEEEOKRLNRT 838

QY 819 -EGISGELKDIHVDHQNLYALTHEGEVFPHPREAWONGAESSSWHKIALPQSESKLSL 877

Db 839 VDRLNKELEQIGDD-----SKLALQQLQAQME-- 865

QY 878 DMSHEHKPIATFEDGSOHQKAGGWHAAYAPERGLAVGTSGSQTFVFNRLMOGVKGKVIP 937

Db 866 ----DYKEKAR-KEVADAQRAQKADWASEAEKNSG-----GLSRLQDELQRLQAL----- 910

QY 938 GSGLTVKLSAQTGMITCAEGRKVSSEKFSERIRAVAFNPTMTSTPRPIKNAAYATGHGQGR 997

```

911 -----QTSQAEEDTARLDKELLAQRLQGLE-----QEA 938
998 EGLKPLYEMOGALIKOLDAHNVHRHNAQPODIQSKLET-LDLGERGAELLND-MKRFREDEL 1055
939 ENKKRFQDDKARQLKSLE-----EKVSRLEAELDEEKNTVELLTDRVNRGRDQV 987
1056 EQSATRSVTVLGQHGVKNGEINSEBKPSPGKALVQSFNVNRSQDLSKSLQQAQVHAT 1115
988 DQRLTELMQERSARQDL-----ECDKISLERQNKDKLTRLASSEGFG 1029
1116 PPSAE-SKLOSMLGHFVSAGVDMSHOKGEIPLGRORPNDKTALTCSRLLDVTYIGELH 1174
1030 KPSASLSQLES-----QNLQERLQABEREKTVLQSTNRKLER-RVKELS 1074
1175 -ELADKAKLVSDHKPDAD-QIKLRQOQDTLRE--KRYESNPVKHYTDMGFTH--NKALE 1228
1075 IQIDDERQHVNQDKQLTLRVKALKRQVDEAEIEIRLDSLRKKAQRELEBEQHEVNEQLQ 1134
1229 ANYDAVKAFINAFKKEHHGVNLTTRTVLESQGSAAELAKKLNLTLLSLD 1276
1135 ARIKSLEK--DAWRK-----ASRSAESA--LKQEGLSDD 1165

```

Search completed: July 7, 2004, 15:18:37  
Job time : 44.0314 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:13:22 ; Search time 81.8129 Seconds  
(without alignments)

7088.403 Million cell updates/sec

Title: US-09-596-784-2

Perfect score: 9448

Sequence: 1 MELKSLGTEHKAHVHTAAHN.....NPQVASALTLKKEGLEMK 1838

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9448	100.0	1838	2 O54581	O54581 erwinia amy
2	5529.5	58.5	1829	2 Q9KH44	Q9KH44 pantoaea agg
3	5502.5	58.2	1835	2 Q9FCY7	Q9FCY7 erwinia ste
4	1831	19.4	1795	16 Q887C9	Q887C9 pseudomonas
5	1824	19.3	1795	2 O66101	O66101 pseudomonas
6	1198.5	12.7	1160	2 Q8RNV8	Q8RNV8 pseudomonas
7	941	10.0	1441	2 Q8VPK4	Q8VPK4 pseudomonas
8	325	3.4	1957	16 Q888Y1	Q888Y1 pseudomonas
9	302.5	3.2	1702	16 Q8XQ66	Q8XQ66 ralstonia s
10	283.5	3.0	9904	16 Q8NMQ6	Q8NMQ6 staphylococ
11	283	3.0	57	2 Q9LAW1	Q9LAW1 erwinia amy
12	275.5	2.9	6713	16 Q9U54	Q9U54 staphylococ
13	275	2.9	2514	16 Q9J30	Q9J30 neisseria m
14	273.5	2.9	6713	16 Q931R6	Q931R6 staphylococ
15	261.5	2.8	1877	4 Q8NEV1	Q8NEV1 homo sapien
16	258.5	2.7	9439	16 Q8CP76	Q8CP76 staphylococ

17	255.5	2.7	3455	16 Q9P9U6	Q9P9U6 xylella fas
18	255	2.7	1875	11 Q8CH77	Q8CH77 mus musculus
19	253.5	2.7	522	2 Q840G7	Q840G7 pseudomonas
20	253	2.7	2045	16 Q9A0K5	Q9A0K5 streptococc
21	251.5	2.7	16223	5 Q8IR22	Q8IR22 drosophila
22	247	2.6	3442	16 Q9PBE8	Q9PBE8 xylella fas
23	246.5	2.6	2813	4 Q8WQ06	Q8WQ06 homo sapien
24	245	2.6	2703	16 Q9X0T0	Q9X0T0 neisseria m
25	243.5	2.6	2813	4 Q96JP6	Q96JP6 homo sapien
26	243.5	2.6	3117	4 Q8WY20	Q8WY20 homo sapien
27	242.5	2.6	3242	2 Q8G3X9	Q8G3X9 escherichia
28	242	2.6	1836	4 Q8IVL2	Q8IVL2 homo sapien
29	241	2.6	2275	16 Q8NUJ3	Q8NUJ3 staphylococ
30	241	2.6	2535	16 Q8ZDR6	Q8ZDR6 yersinia pe
31	241	2.6	2579	16 Q8DOR8	Q8DOR8 yersinia pe
32	240.5	2.5	1906	11 Q8OTI7	Q8OTI7 mus musculu
33	240.5	2.5	5327	5 Q76891	Q76891 drosophila
34	239	2.5	2712	16 Q9F3X5	Q9F3X5 pasteurella
35	238	2.5	2283	2 Q8VQ99	Q8VQ99 staphylococ
36	238	2.5	4042	2 Q8KT64	Q8KT64 photorhabdu
37	237.5	2.5	2817	4 Q96P79	Q96P79 homo sapien
38	237.5	2.5	4152	2 Q9ZHL3	Q9ZHL3 haemophilus
39	237.5	2.5	4152	16 Q7VLE8	Q7VLE8 haemophilus
40	236.5	2.5	3012	5 Q97205	Q97205 leishmania
41	236	2.5	2271	16 Q99QY4	Q99QY4 staphylococ
42	236	2.5	3241	2 Q9AHF9	Q9AHF9 fusobacteri
43	235.5	2.5	3222	16 Q8XQZ5	Q8XQZ5 ralstonia s
44	231	2.4	3692	16 Q8CNU9	Q8CNU9 staphylococ
45	230.5	2.4	5020	16 Q8E9W3	Q8E9W3 shewanella

## ALIGNMENTS

### RESULT 1

ID	O54581	PRELIMINARY;	PRT;	1838 AA.
AC	O54581;			
DT	01-JUN-1998	(TrEMBLrel. 06, Created)		
DT	01-JUN-1998	(TrEMBLrel. 06, Last sequence update)		
DT	01-JUN-2003	(TrEMBLrel. 24, Last annotation update)		
DE	DSPE.			
GN	DSPE OR DSPA.			
OS	Erwinia amylovora.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Erwinia.			
OX	NCBI_TaxID=552;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EA321;			
RX	MEDLINE=98115919; PubMed=9448330;			
RA	Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,			
RA	Conlin A.K., Collier A., Beer S.V.;			
RT	"Homology and functional similarity of an hrp-linked pathogenicity			
RT	locus, dspeP, of Erwinia amylovora and the avirulence locus avrE of			
RT	Pseudomonas syringae pathovar tomato."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CFBP1430;			
RX	MEDLINE=98086111; PubMed=9426142;			
RA	Gaudriault S., Malandrin L., Paulin J.F., Barny M.A.;			
RA	"DspA, an essential pathogenicity factor of Erwinia amylovora showing			
RT	homology with AvrE of Pseudomonas syringae, is secreted via the Hrp			
RT	secretion pathway in a DspB-dependent way."			
RL	Mol. Microbiol. 26:1057-1069 (1997).			
RN	[4]			
RP	SEQUENCE OF 1-238 FROM N.A.			

RC STRAIN=EA321, and ATCC 49947;  
 RX MEDLINE=98422475; PubMed=9748455;  
 RA Kim J.F., Beer S.V.;  
 RT "HrpW of *Erwinia amylovora*, a new harpin that contains a domain  
 homologous to pectate lyases of a distinct class.";  
 RL J. Bacteriol. 180:5203-5210 (1998).  
 DR EMBL; U97504; AAC04850.1; -;  
 DR EMBL; Y13831; CAA74156.1; -;  
 DR EMBL; U94513; AAC62315.1; -;  
 DR PIR; T18448; T18448.  
 SQ SEQUENCE 1838 AA; 198243 MW; 7D595F78130E8FD9 CRC64;

Query Match 100.0%; Score 9448; DB 2; Length 1838;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSPQNAASLAAGKNGKMPRIHQ 60  
 DB 1 MELKSLGTEHKAHVTAANPVGHVALOOGSSSSPQNAASLAAGKNGKMPRIHQ 60

QY 61 STAADGISAHQOKKFSLRGCLGTTKFSRPAQGGTTHSKGATRLDARDGCTQH 120  
 DB 61 STAADGISAHQOKKFSLRGCLGTTKFSRPAQGGTTHSKGATRLDARDGCTQH 120

QY 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVYKGGSGEDKVTQQRHQLNFFGQROTMLS 180  
 DB 121 EAAAPDAARLTRSGGVKRRNMDMAGRPVYKGGSGEDKVTQQRHQLNFFGQROTMLS 180

QY 181 KMAHPASANAGDRLOHSPHPIPGSHHEIKKEEPPVSTSKATTADHADVETAEQEDDSEFOQ 240  
 DB 181 KMAHPASANAGDRLOHSPHPIPGSHHEIKKEEPPVSTSKATTADHADVETAEQEDDSEFOQ 240

QY 241 LHQQLARENPQPQKLGVAFTPI SARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300  
 DB 241 LHQQLARENPQPQKLGVAFTPI SARFQPKLTAVAESVLEGTDTTQSPKLPQSMKGS 300

QY 301 GAGVTPPLAVTLKQKQLAPDNPALNTLLKQTLGKDTQHYLAHASSDQSQHLLDNKG 360  
 DB 301 GAGVTPPLAVTLKQKQLAPDNPALNTLLKQTLGKDTQHYLAHASSDQSQHLLDNKG 360

QY 361 HLFDIKSTATSVLHNSHPGETKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420  
 DB 361 HLFDIKSTATSVLHNSHPGETKGLAQAGTGSVSDGKSGKISLGSGTQSHNKTMLSQ 420

QY 421 PGEAHSRLTGTWQHAPAGAPGESIRLHDDKIHLHPGLGWQSDADKTHSLSRQAD 480  
 DB 421 PGEAHSRLTGTWQHAPAGAPGESIRLHDDKIHLHPGLGWQSDADKTHSLSRQAD 480

QY 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQGVQVAILTDTGPRHKMSIMPSLDAS 540  
 DB 481 GKLYALKDNRTLQNLSDNKSSEKLVDKIKSYSDQGVQVAILTDTGPRHKMSIMPSLDAS 540

QY 541 PESHISLSLHFAHDAHQGLLHGKSEBAQSVASHGRLVVDSSGKLFSAAI PKQGDGNEL 600  
 DB 541 PESHISLSLHFAHDAHQGLLHGKSEBAQSVASHGRLVVDSSGKLFSAAI PKQGDGNEL 600

QY 601 KMKAMPQHALDEHFGHDHQSIFPHDDHQLNALVKNFRQQAHCPLGNDHPQGNL 660  
 DB 601 KMKAMPQHALDEHFGHDHQSIFPHDDHQLNALVKNFRQQAHCPLGNDHPQGNL 660

QY 661 DALVIDNQLGHTNPEPHEIILDMGHLSLALQEGKHYFDQTLTKGWTGAESDCKQLKKG 720  
 DB 661 DALVIDNQLGHTNPEPHEIILDMGHLSLALQEGKHYFDQTLTKGWTGAESDCKQLKKG 720

QY 721 LDGAAVLLKDGGEVKRLNIQSTSSIKHGTENVFSLPHVRNKPDPGALQGLNKDKQAQAM 780  
 DB 721 LDGAAVLLKDGGEVKRLNIQSTSSIKHGTENVFSLPHVRNKPDPGALQGLNKDKQAQAM 780

QY 781 AVIGVKNYALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIIHVDHKQNIYALT 840  
 DB 781 AVIGVKNYALTEKGDIRSFQIKPGTQQLERPAQTLSREGISGELKDIIHVDHKQNIYALT 840

QY 841 HEDEVPHQPREAWQNGAESSSWHKLALPQSESKLSLDMSHEHKPTATFEDGSQHQLKAG 900

DB 841 HEDEVPHQPREAWQNGAESSSWHKLALPQSESKLSLDMSHEHKPTATFEDGSQHQLKAG 900  
 QY 901 GHIAVAAAPBPGPLAVGTSGQVFNRLMGVKGKVI PGSGLTVKLSAQGTGMTGABGRKV 960  
 DB 901 GHIAVAAAPBPGPLAVGTSGQVFNRLMGVKGKVI PGSGLTVKLSAQGTGMTGABGRKV 960

QY 961 SSKFSERIRAYAFNPMTSTPRPIKNAAYATQHGQWQREGKLPLEYOGALIKOLDAHNVR 1020  
 DB 961 SSKFSERIRAYAFNPMTSTPRPIKNAAYATQHGQWQREGKLPLEYOGALIKOLDAHNVR 1020

QY 1021 HNAPODLOSLETLDLGEHGAELNDMKRPFDELEQSATRSVTVLGOQGVLSKNGEIN 1080  
 DB 1021 HNAPODLOSLETLDLGEHGAELNDMKRPFDELEQSATRSVTVLGOQGVLSKNGEIN 1080

QY 1081 SEFKSPGKALVQSFNVRNSGQDLKSLSQAAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140  
 DB 1081 SEFKSPGKALVQSFNVRNSGQDLKSLSQAAVHATPPSAESKLSQMLGHFVSAGVDMSHQ 1140

QY 1141 KGEIPLGRORDNDKTALTKSRLIILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1200  
 DB 1141 KGEIPLGRORDNDKTALTKSRLIILDTVTIGELHELADKAKLVSDHKPDADQIKQLRQOF 1200

QY 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTRTVLESQ 1260  
 DB 1201 DTLREKRYESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTRTVLESQ 1260

QY 1261 SAEALAKKNTLLSLDSESMFSRSYGGGVSTVFVPTLSKKVPVPIVPIGAGITLDRAYN 1320  
 DB 1261 SAEALAKKNTLLSLDSESMFSRSYGGGVSTVFVPTLSKKVPVPIVPIGAGITLDRAYN 1320

QY 1321 LSFRTSGGLNVSFGRDGGVGNIMVATGHDVWPMYMTGKTSAGNASDMLSAKHKISPD 1380  
 DB 1321 LSFRTSGGLNVSFGRDGGVGNIMVATGHDVWPMYMTGKTSAGNASDMLSAKHKISPD 1380

QY 1381 RIGAAVSGTLQGTQNLKFKLTDELPGFIHGLTHGTLTPAELLQKGIHQHMQGSKLT 1440  
 DB 1381 RIGAAVSGTLQGTQNLKFKLTDELPGFIHGLTHGTLTPAELLQKGIHQHMQGSKLT 1440

QY 1441 FSVDTSANLDRAGINLINEDGSKPNGVTARVSAGLSASANLAAGSERSTTSQFGSTTS 1500  
 DB 1441 FSVDTSANLDRAGINLINEDGSKPNGVTARVSAGLSASANLAAGSERSTTSQFGSTTS 1500

QY 1501 ASNNRPTFLNGVAGANLTAALGVHSSTHEKPVGIPPAFTSTNVSAALDNRTSQSI 1560  
 DB 1501 ASNNRPTFLNGVAGANLTAALGVHSSTHEKPVGIPPAFTSTNVSAALDNRTSQSI 1560

QY 1561 SLELKRAEPVTNDISELTSTLGKHPKDSATTMKAALKELDDAKPAEQHLILQHFSAK 1620  
 DB 1561 SLELKRAEPVTNDISELTSTLGKHPKDSATTMKAALKELDDAKPAEQHLILQHFSAK 1620

QY 1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGASHSSTTYNNLSRINNDGI VELLKHFD 1680  
 DB 1621 DVVGDERYEAVERNKLKLVIRQQAADSHSMELGASHSSTTYNNLSRINNDGI VELLKHFD 1680

QY 1681 AALPASSAKRLGEMMNDPALXDI IKQLOSTPFSSASVSMELKDGRLQTEKAILDGKVG 1740  
 DB 1681 AALPASSAKRLGEMMNDPALXDI IKQLOSTPFSSASVSMELKDGRLQTEKAILDGKVG 1740

QY 1741 REEVGVLFODRNLNRKVSQSVSKSGENTPALLGTSNSAAMSMERNITGTFNKYG 1800  
 DB 1741 REEVGVLFODRNLNRKVSQSVSKSGENTPALLGTSNSAAMSMERNITGTFNKYG 1800

QY 1801 QDQNTFRRTLEGGIAQANPOVASALTDLKKEGLEMS 1838  
 DB 1801 QDQNTFRRTLEGGIAQANPOVASALTDLKKEGLEMS 1838

RESULT 2  
 Q9XH44  
 ID Q9KH44  
 AC Q9KH44;  
 DT 01-OCT-2000 (T-EMBLrel. 15, Created)



```
QY 1835 EMKS 1838
Db 1826 EMKS 1829

RESULT 3
Q9FCY7 PRELIMINARY; PRT; 1835 AA.
AC Q9FCV7
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE WSE.
GN WSE.
OS Erwinia stewartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=66271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS104;
RA Frederic R.D., Ahmad M., Majerczak D.R., Arroyo-Rodriguez A.S.,
RA Manulis S., Coplin D.B.;
RT "Genetic organization of the Pantoea stewartii subsp. stewartii hrp
RT gene cluster and sequence analysis of the hrpA, hrpC, hrpN and wse
RT operons.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF282857; AG01467.1; -
SQ SEQUENCE 1835 AA; 201363 MW; 546738CB5C151A4E CRC64;

Query Match 58.2%; Score 5502.5; DB 2; Length 1835;
Best Local Similarity 58.8%; Pred. No. 1.5e-274;
Matches 1095; Conservative 269; Mismatches 446; Indels 53; Gaps 15;

QY 1 MELKSLGTEHKAHVTHAHPVGHV-ALQCGSSSPQNAASLAAGKRGKMPRIHQ 59
Db 1 MRLNLHATEKKTIVQV-VENPNNSTIPPLQCGSSSAPQASGGTLAGSEGNIAASMPAIRQ 59
QY 60 PSTAADGISAHQKSFSLRGCLGTGKFKRSAPQ-----GPGTTHSGATRLDILLAR 113
Db 60 HLSADGNGAARQKNSFSLKFNLFGLFKTKTKARAQPTNPGSNVPESRIIRPLGDFLAQ 119
QY 114 DDGETQHEAAPDAA--RLTRSGVVRNMDMDWAGRPVMVGGGDEKVPYQOKRHQLNNF 171
Db 120 PEQDGEIQAPHAASAPRLTRSPGVSRHSVEDNRDKPVAKADPRE--TSAIDIKHQLNNF 177
QY 172 GQNRQTMWSK---MAHPASANAGDRLOHSPHPHPSHHEIKERPVGSTSKATTAHADRVE 228
Db 178 NOMRONILNTRPFSQGATTSVMDARAPLPAIPSTTSEITEE-----VD 222
QY 229 IAEDEDDSEFQOLHQORLARENPPOPPKLGVAATPISARFQPKLTAVASVLEGTDTQ 288
Db 223 ATQPE-----QQLPORHL-DARENPSILKTRDVTSHIRFRPTLPAIENTAEANRR 276
QY 289 -SPLKPSOMLKGSGAGVTPPLAVTLDDKGLQOLAPNPPALNTLLKQTKGKTOHYLAHAS 347
Db 277 IQPLQTPPPLENT--PVSPLSLTLDKGLRLADSNPTAMNTLLQTLGKEGQHYVAHTAS 334
QY 348 SDGSQHLHLLDNKGLFDIKSTATSYSLVHNSHPEIKGLAQAGTGSVSDGSKGISLG 407
Db 335 AEGNHMMLDEQGHFLFKLHSHGEGYSVFHNSLPSTVRQMSKNNDPEIKLVNDNGKLILD 394
QY 408 SGTQSHNKTMLSOPGEAHRSLITGIWQHP-AGARPQGESIRLHDDKIHLHPELGVWQS 466
Db 395 RGAQGEVHLNPPGEIHAYLSGIVRHSVAGENNQWVRIHDDKIHYLNNEVGWQT 454
QY 467 ADKTHQSLQRADKGLYALKDNKRTLQNLSDNKSSEKLVKIKSYVDQRGQVAILDTTP 526
Db 455 SDKTPYQSLVQSGDKLYAYRDSRALHNLSDNHVSFVDKIKSFVSEHGEVAULTDTE 514
QY 527 GRHKMSTMPSLDASPESHISLSLHFADAHQGLLHGKSELEAQSVAISHGLRVADSEGL 586
Db 515 SPFHICLMPSANASSERIAESLHLDATMQLQRBPHLETQSTALDKDGRLLFAADSEGL 574
```

```
QY 1656 HSTYNNISRNNDGIVELLHHKHFDAALPASSAKELGEMMNNDPALKDIIKOLQSTPSS 1715
D 1653 HTTQYSLRLDENGIFILHHFNAALPPTSATRISNMANNPLKALIQKQSSPTS 1712
QY 1716 ASVSMELDGLREOTEKAILDQKGVREEVGLFODRNLRVKSVSOSVSKSEGFNTPA 1775
D 1713 ARVSMELDGLRDQTEKAILDQKGVREELARFODRNLRIRSINVSOSVSKSEGFNPT 1772
QY 1776 LLLGTSNAAMSMEINIGTINFKYQDQDTPRRFTLEGIGIAQANPQVASALTDLKEGLE 1835
D 1773 SILGASNSAGLSMGRNIGTINFKYQDQDTPRRFTLEGIELAKANPDVASALSLEKGEF 1832
QY 1836 MKS 1838
D 1833 MKS 1835

RESULT 4
ID Q887C9 PRELIMINARY; PRT; 1795 AA.
AC Q887C9;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE Avirulence protein Avre (Pto).
GN AVRE (Pto) OR PSPT01377.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Uterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016860; AA054899.1; -.
DR TIGR; PSPT01377; -.
KW Complete proteome.
SQ SEQUENCE 1795 AA; 6C9BF0D71364A08C CRC64;

Query Match
Best Local Similarity 29.5%; Pred No. 2.8e-85;
Matches 564; Conservative 363; Mismatches 737; Indels 248; Gaps 75;

QY 55 PRIH-----QPSTAADGISAH-----QOKKFSFLRGCLGT-----KKFSRS- 91
D 4 PSIHRTGSIITPTVTPDARAATDLQERAEQPRQRSSHSL--SSVGKRAKSVGKLFQSK 62
QY 92 APOQOPGT-----THSKATLDDLARDG-----ETQHEAAPPDARLTRSGV 136
D 63 APOQKAATPTTAKNVKTPPPASNVATPRN-KARESGFSNSPQNTHSAPKSI----- 113
QY 137 KRRNMDMAGRPVMYKGGSGEDKVPPTQQRHOLNFGQWQMTLSKMAHPASANAGRLQH 196
D 114 -----LRNHPNQAASSG-----AQTHEIPEAPRKNLRVRFDLQ-----DRUER 154
QY 197 SPHPHPSHHEKEEPVGSTSKATTADRVIEIAQEDDDSEFQQLHQORLARENPPOP 256
D 155 SPSYLDSDNPMTDEAVANATRF-----RSPDSLQSGDGRISMLATDPDPQ 203
QY 257 ----PKLGVAAT-PISAROPKLTAVAESVLGTDTTQPLKPKQSMKSGAGVTPL-AVT 310
D 204 SSSGSKIQSDSGPIPR-----EPMLRWSNGRFPKDEKLVRNS-----EPQSGIQ 250
QY 311 LD-KGKLQAPDNPALNTLLKQTLGKDTQHYLAHASSDGSQHLLLDNKGHLFDIKSTA 369
D 251 LDAGKPDFTFTPTGLAPLDSILATPKQTYLAHQSKDGVGHQLQANGHLLHLAQDD 310
```

```
QY 370 TSYSLYHNSHFG-EIKGKLAQAGTGSVSDGKSGKISLGSQTSQSHNKTMLSQFBAHRS 428
D 311 SSLAVIRSSNEALLIEGKPPA---VKMEREDGNIHI--DTASGRKTQ-ELPGKAHIAH 363
QY 429 LTGTWQHPAGAAPQGESIRLHDKIHLHPELGWQASD--KDT-HSOLSHQADOKLYA 485
D 364 ITNV-----LLSHDGERMYRVEDRLYQFDPISTRWKIPEGLEDTPAPNSLSTGNGSVYA 417
QY 486 LKDNRTTLONLSDNKSSEKLVKIRKSYVDQGOVAILTDTQGRHKWSIMPSLSDASPESHI 545
D 418 KSDDAV--DLSPMPWPHVEEDLQSFVAPNRAALLS---GKTQAILLT-DMGP----- 468
QY 546 SLSLHFADAHQGLLHGKSELAQSAVISHGLRVADSEGLFSAAIKPGQDGNELMKAM 605
D 469 VIGGLTPKTKGLELDGGAQAAGVLSGDKLFIADTQGRLYSAD-RSAFEGDDPKLKL 527
QY 606 POHALDE---HFQHDHQSIFPFDHDDHQLNALVKNFRQOHCAPLG-NDHQFHPGWNLT 660
D 528 PEQANFQLEGVPLGCHNRVTGFIINGDDGGVHALIKNRQETHSHALDEQSSKLOSGWNL 587
QY 661 DALVIDNQLGLHHTNPPEP---HEILDHGLSLAQEGKLHVFQDLTKGWTGAE--SDCKQ 716
D 588 NALVLNNRGL--TWPPPTAADRLNLDRAGLVGLSEGRIOKWDATPCWKDAGIKDIDR 645
QY 717 LKGLDGAAYLLKXGDEVKVRKLNINQSTSIKHGTENVFSLPHVRNKPPEPDALQGLNKDK 776
D 646 LQRGADSNAYVLKGGKLHALKIAAEHPNMAFDRTALTAQTAARSTKVMGKEIEGL--DDR 703
QY 777 A-QAMAVIGVKNYLALTEKGDIRSFQIKPGTQQLERPAQTLSREGISELGDHIVDHQON 835
D 704 VIKAFAMVSNKRFVALDD-----QNKLTAKSHDKHPVTLDIFGLEGDIKLSLDEKHN 756
QY 836 LYALTHTHEGHPQPREAWON---GAE-SSSWHKLALPOSESKLSLSDMSHSHKPTATED 891
D 757 LHALTSTGLYCLPKEAWQSTKLGDQLPARWTPVALPGQ--PVKALFTDNDNVLSAQIED 815
QY 892 GSQH---QLKAGGWHAYAAPERGLAVTSGTSQTVFNRLMQGVKGVKIPGSLGVTKLSAQ 948
D 816 AEGKGLMQLKAGQWRP---EORP--VEENGLNDVHSRIITGNSKNTWRIPKTLGLTLMVDN 870
QY 949 TGMGTGAEGRKVSSEKFSERIPAYAFNPTMSTPRIPKNAAYATQHWGQREGKPLPYEMQG 1008
D 871 TFGRSQVGEKSKAST-SEFIRANTYKNTAETPRWKNVGVGHQHYQGRGLKVEYETES 929
QY 1009 ALIKOLDAHNVNRNAPOP-----DLQSKLETLD--LGEHGAELNDMKRPREDELEQSATR 1061
D 930 MLFKQLEL--IHESGGPPAPGQDLKARITALEAKLGPQGAATLVEKLETLDELENHSYT 987
QY 1062 SVTVILG-----OHQGVLSNGEINSFKEPSP---GKALVQ---SFNVNRSGQDL 1104
D 988 ALMSIGQSYGKAKNLKQDDGILNQHGEL---AKFSVRMQFGKKLADLTGKLNFKSSGHDL 1044
QY 1105 SKSLQQAATHATPPSAESKLSMLGHFVSAGVDMHSHQGEIPLGRQRPNDKTAITKSLRI 1164
D 1045 VKELQDALTOVAPSAENPTKLLGLTKHQGLKLSHQKADIPLGQRDRASEDHGLSKARLA 1104
QY 1165 LDTVTIGELHELADAKLVSDHKPDQDIKOLRQOPDTLREKRVESNPVHYTDMGFTHN 1224
D 1105 LDLVTLSLGLALDQV-----EQLPQSDIPLQKKLATLRTVYGENPVKVVDMGFTHN 1160
QY 1225 KALEANYDAVKAFINAFKKEHHGVNLTTRTVLESQGAELAKKLKNTILSLDSG-ESMSF 1283
D 1161 KALESGYESVKTFLKSPKKADHAVSVNMRATGSKDOALAGKPKMLKQLEHGDDEVL 1220
QY 1284 SRSYGGGVSTVFPVPTLSKKVPVPIPGAGITLDRAYNLSFSRTSGGLN---VSFGRGGV 1340
D 1221 QRSYGVNLTPTPFIILADK--ATGLWPTAGATGNRYITLNAERCEGVTVLVISEGA-GNV 1277
QY 1341 SGNIMVATGHDVWPEYM-----TGKKTSGAGNASDWLSAKHKLSPLDRIGAAVSGTLQGT 1395
D 1278 SGGF--GAGKDYWPGFFDANNPAKSVVDGN-----NRTLTPNFRGLVDVTATVAASQR 1328
QY 1396 NSLKFKLTEDELPGFIHGLTHGTLTPAELQKGIHQKQSKLTFSDVTSANLDRAGI 1455
```

Db 1329 AGVVNVPEBIDDAFVDDLFEGQLNPLQVLKAVDHSVEARFNFDTAGGTADIRAGI 1388  
 QY 1456 NLNED---GSKPN--GVTARVSAGLSA--SANAAGSRERSTTSGQFGSTTSASNRPTF 1508  
 Db 1389 NLTEDRDPNADPNSDSFSVAVRGGAANIITVNIIMTYDYSLTQKNDKTELKGGKRNPRF 1448  
 QY 1509 LNVGVGAGANLTAALGVAHSSHTGKPV--GIFPA--FTSTNVSAAL--ALDNRITSOSIS 1561  
 Db 1449 LNNVTAGGQRAIGSGHTAP--TGTASAPGPTPASCATAANNLGGALNFSVENVTKRIK 1507  
 QY 1562 LELKRAPVTSNDISELTSTLGHFKDSDATTKMLAALKELODA-----KPAE-----QLHI 1612  
 Db 1508 FRYNVAKPIITTEGLSKSLGKEAFDNTTKAKLAELADPLNARYTGGKPDVIAQLDG 1567  
 QY 1613 LQOHFS---AKDVGVDERVEARNLKLVIROQA--ADSHSMELGSAHSTTYNNLSRIN 1667  
 Db 1568 LEELFADIPPKD--NDKQYKALRDLKRAAVERASANKHSV--MDNARFETSNTNLSGLS 1624  
 QY 1668 NDGIV-ELLKHFDPAALPASSAKRLGEMMNNDPALKDIIKQLQSTPPSSASVSMELKDGL 1726  
 Db 1625 SESLTKIMSSVRDASAP--GNATRVAFEMRQDPKLRAMLKEMSGSIGTLARVLEPKUSL 1683  
 QY 1727 REOTEKAILDKVGRBEVGVLFQDRNNLRKVSVSQSVKSGGFNTPTPALLIGTSNAAM 1786  
 Db 1684 VDKIDGSLNGTMTQSDLSMLEDRENMRIKRLVVFHTATQAEFTTPTPLVSNSGANV 1743  
 QY 1787 SMERNITGTFNKYQOQNTPRRTLEGGLAQANPOVASALTDILKKEGLEMK 1838  
 Db 1744 SVTKTLGRINVFYGAQDDKPIGYTFDGLSRPSASLKEAAGDLKSGFELKS 1795

## RESULT 5

O66101 PRELIMINARY; PRT; 1795 AA.  
 AC O66101;  
 DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
 DE Avirulence protein (AVRE).  
 GN AVRE.  
 OS Pseudomonas syringae (pv. tomato).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 OX NCBI\_TaxID=323;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=98115919; PubMed=9448330;  
 RA Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,  
 RA Conlin A.K., Collmer A., Beer S.V.;  
 RT "Homology and functional similarity of an hrp-linked pathogenicity  
 locus, dmpE, of Erwinia amylovora and the avirulence locus avrE of  
 Pseudomonas syringae pathovar tomato.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=98422476; PubMed=9748456;  
 RA Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,  
 RA Collmer A.;  
 RT "The Pseudomonas syringae pv. tomato HrpW protein has domains similar  
 to harpins and pectate lyases and can elicit the plant hypersensitive  
 response and bind to pectate.";  
 RL J. Bacteriol. 180:5211-5217 (1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=DC3000;  
 RX MEDLINE=20243785; PubMed=10781092;  
 RA Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,  
 RA Pernicki-Ocwleja T., van Dijk K., Collmer A.;  
 RT "The Pseudomonas syringae Hrp pathogenicity island has a tripartite  
 mosaic structure composed of a cluster of type III secretion genes

RT bounded by exchangeable effector and conserved effector loci that  
 RT contribute to parasitic fitness and pathogenicity in plants.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861 (2000).  
 DR EMBL; U97505; AAC06134.1; -.  
 DR EMBL; AF232004; AAF71499.1; -.  
 DR PIR; T30332; T30332.  
 SQ SEQUENCE 1795 AA; 195351 MW; 260F74534DE08D5F CRC64;

Query Match 19.3%; Score 1824; DB 2; Length 1795;  
 Best Local Similarity 29.0%; Pred. No. 6.5e-85;  
 Matches 555; Conservative 366; Mismatches 744; Indels 246; Gaps 72;

QY 55 PRIH-----QBSTAADGISAHQKKSFSRLRGCLGTGKFSRSAPQCGTTHSKCA-- 105  
 Db 4 PSIHRTGSIQPTVTPDARAATDLOER-----AEQPRSSHSLSVSG 47  
 QY 106 --TLRDL--LARDGETQHEAAAPDAAR-----LTRSGGVKRNMD-- 143  
 Db 48 KRALKSVGKLFQSKAPQQAATPTAKNVKTPPPASNVATPRNKAREGFSNSPONT 107  
 QY 144 -----MAGRPVKGSGEDKVPETQOKRHQNLNFGQMRQTMLSKMAHPASANAGDRLQHS 197  
 Db 108 RAPKWILRNHPNQASSG-----AQTHEIPEAAPRKNLRVRFDLQ-----DRLERS 155  
 QY 198 PPHIPGSHHEIKKEPVGSTSKATTADRVDAVEIAQEDDDSEFQHLHQORLARENPPOP- 256  
 Db 156 PSYLDSDNPMTDEEAVANATRF-----RSPDSHLOGSDGTRISMLATDPDPS 204  
 QY 257 ---PKLGAT--PISARFQPKLTAVESVLEGTDTTOSLPKPSQMLKSGAGVPL-AVTL 311  
 Db 205 SSGSKIGSDGPIPPR-----EPMLWRSGRPFELKDEKLVRNS-----EPQSGIQL 251  
 QY 312 D-KGKLQALPADNPALNTLLKQTLGKDTQHYLAHASSSDGSQHLNLLDNKHLFDIKSTAT 370  
 Db 252 DAKGKPDFSTFTPLGLPLDLSILATPKQYLAHQSKDGVGHQLQANGHFLHLAQDSS 311  
 QY 371 SYVLNHNHPG-EIKGKLAQAGTGSVDGSKSKISLGSTQSHNKTMLSQPGEAHRSLL 429  
 Db 312 SLAVIRSSNEALLIEGKKPPA---VKVEREDGNIHI--DTASGRKTQ-ELPGKAHTAHI 364  
 QY 430 TGIWQHPAGAARPOGESIRLHDDKIHLHPGLGWQSD--KDT-HSOLSRQADKLYAL 486  
 Db 365 TNV-----LLSHDGERMRVHEDRLVQFDPISTRWKIPEGEDTAFNSLTGGNGSVYAK 418  
 QY 487 KDNRTLQNLSDNKSSEKLVDKISYSVDQGVQVAILTDTPCRHKWSIMPDLSDASPEHS 546  
 Db 419 SDDAVV-DLSSPFMPHVEVEDLQSFVAPDNRAALLS---GKTQAILLT-DMSP---V 469  
 QY 547 LSLHFDADAHQGLHKGSELEAQSVAISHGLVVDSEKLFSAAPKQGDGNELMKAMP 606  
 Db 470 IGGITPKTKGLELDGGKQAQAAVGLSGDKLFIADTQGRLYSAD-RSAFEGDDPKLIMP 528  
 QY 607 QHALDE---HFGHDHIOISGFFHDDHGQLNALVKNFRQOQHACPLG-NDHQFHPGNLTD 661  
 Db 529 EQANFQLEGVPLGHNRTVTFINGDDGVHVALIKNRQGETHSHALDQSSKLSQSGNLTN 588  
 QY 662 ALVIDNQLGLHHTNPEP---HEILDGMHGLSLAQEGKJHYFDQLTKGWTGAE--SDCKQL 717  
 Db 589 ALVLNNRGL--TWPPPTAADRNLNDRAGLVGLSEGRIOQRWDATPECWKDAGIKDRL 646  
 QY 718 KGLDGAAYLLKQGEVVRKLNINQSTSIKHGTENVFSLPHVRNKPEGDALQGLNKDKA 777  
 Db 647 QRGADSNAYVLKGGKHALKIAAEHPNMAFDRNTALAQATARSTKVMKEIEGL--DORV 704  
 QY 778 -QAMAVIGVKNKLYALTEKGDIRSFQIKPGTQQLERPAQTLRSREGISGELKDIDHVDKQNL 836  
 Db 705 IKAFAMVSNKRFVALDD-----QNKLTAKSKDHPVTLTDIPGLEGDIKSLSDKXNL 757  
 QY 837 YALTHEGEVHPQREAWON---GAE-SSSWHKALPOSESKLKLDSMHSHEHKPTATFEDG 892  
 Db 758 HALTSTGGLYCLPKPEAWQSTKLGDLQRLARWTFVALPGQ--PVKALFTNDNDVLQAIEDA 816  
 QY 893 SQH---QLKAGGWHAYAAPERGPLAVGTSGSQTFVFNLMQGVKGKVI PGSGLTVKLSAQ 949

Db 817 EKGLMLQKAGOWQF---EORP--VEENGLNDVHSRITGSGNKTWRIKFTGLTLMEDVNT 871  
Qy 950 GGMTAGRKVSSKFSRIRAYFNPWTSTPRIPKAAVATQHGQWQEGELKPLVEMOGA 1009  
Db 872 FGRSGVEKSKAST--SEFIRANIYKNTAETPRWKNVGDHQRHYQGRGLGLKEVTVETSM 930  
Qy 1010 LKQLDANVRHNPAP-----DIQSKLETLD--LGEHGAELNDMKRFRDELEQSATRS 1062  
Db 931 LFKQLEL--IHESGRPPARGQDLKALFAELKLPQCATLVKELETIRLELHNSVTA 988  
Qy 1063 VTVLGL-----QHOGVLSKNSCEINSEKPPSP-----GKALVQ---SPFNVRSGQDLS 1105  
Db 989 LMSICQSGYKAKNLKQDGIINLQHGEL---AKPSVRMQFGKKLADLGTAKMFKSSGHDVL 1045  
Qy 1106 KSLQAVHATPPSABSKLSQKLGHPVSAGVDMGQGEIPLGRQRPDNDKATLTKSRLLIL 1165  
Db 1046 KELQDALTVQAPSAENPTKLLGTGLKHQGLKLSHQKADIPLQRRDASEDHGLSARIAL 1105  
Qy 1166 DTVTIGELHELADKAKLVSDHKPDADQIKQLRQQPDTTLREKRYESNPVKHYTMDGFTHNK 1225  
Db 1106 DLVTLKSLGALLDQV---EQLPQSDIETPLQKLLATLRDVTYGENPVKVVTDMGFTDNK 1161  
Qy 1226 ALEANYDAVKAFINAFKKEHGVNLTTRTVLESQGSABELAKKLNTLLSLDSG--BSMSPS 1284  
Db 1162 ALESYGVSVKTLFSFKADHAVSNMRAATGSKDQAEAGLQKLEHGDDEVLQ 1221  
Qy 1285 RSYGGGVSTVFTLSKKVPVPIPGAGITLDRAVNLFSRSTSGLN--VSFGRDGGVS 1341  
Db 1222 RSYGNLTPFILLADK--ATGLPFTAGTGRNVLNARECEGGVTLYLISEGA--GNVS 1278  
Qy 1342 QNIMVATGHDVWYV-----TGKTSAGNASDWLSAKHKISPDLRIGAASVTGLQTLQN 1396  
Db 1279 GGF--GAGKDYWPFGFDANPARSDVGN-----NRTLTPNRLGVDVTATVAASQRA 1329  
Qy 1397 SLXKELTEDELPGRIHGLTHGLTTPAEILLQKGIHQMKQSGKLTPSVTNSANLDRAGIN 1456  
Db 1330 GWFNVNPDIEDAFVDDLFEQOLNPLQVLKXAVDHESEYARRFNFDLTAGGTADIRAGIN 1389  
Qy 1457 LNEED---GSKPN--GVTVARVSAGLSA--SANLAAGSRERSSTTSQFGSTTSASNNRPTFL 1509  
Db 1390 LTERDPNADPNSSDSFSVAVRGFGAANTVNLMTYDLSLTKQKNDKTELKGGKNNRPRFL 1449  
Qy 1510 NGVAGANLTAALGVAHSTHEGKPV---GIFPA--FTSTNVSAAL--ALDNRTSQSISL 1562  
Db 1450 NNVTAGGQLRAQIGGSTAP--TGTPASAPGTPASQTAANNLGGALNFSVENRTVKIKF 1508  
Qy 1563 ELKRAEPTVNDISLSTLSTLSTLSTLSTLSTLSTLSTLSTLSTLSTLSTLSTLSTLSTL 1613  
Db 1509 RYNTAKPITTEGLSKSLGSLGEAFDNTTTRAKLAELADPLNARYTGKPKDEVIQAOLDGL 1568  
Qy 1614 QOHFS---AKDVVDGDERYEAVRNLKLVIRQQA--ADSHSMELGSAHSSTYNNLIRINN 1668  
Db 1569 BELFADIPPPKD--NDKQYKALRDLKRAVBRHSANXHSV--MDNARETSTKNLSGLSS 1625  
Qy 1669 DGIV--ELLKHFDALPASSAKRIGEMNNPDALKDIIKQLQSTPFSASVSMELKQGLR 1727  
Db 1626 ESILTKIMSSVRDASAP--GNATRVAEFMRQDPKRLAKMEKESIGTLARVLRPKDSIV 1684  
Qy 1728 EQTEKAILDGKVGREEVGLFQDRNNLRVKSVSQSVSKSEGENTPALLIGTNSAAMS 1787  
Db 1685 DKIDEGSLNGTQSDLSMLEDREMEIKLVVPHFTATQAEFTSTPTPLVSYSNGANVS 1744  
Qy 1788 MERNIGTINFKYGODNTPRFTLEGGIAQANPQVASALTDLKKEGLMKS 1838  
Db 1745 VTKILGRINFYGADQDQKPIGYTDFGELSRFSASLKEAAGDLKKEGFLKS 1795

RESULT 6

Q8RNY8

ID Q8RNY8 PRELIMINARY; PRT; 1160 AA.

AC Q8RNY8

DT 01-JUN-2002 (TremBLrel. 21, Created)

DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE Type III effector AvrEpa (Fragment).  
OS Pseudomonas syringae pv. maculicola.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=59511;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ES4326;  
EX MEDLINE=21862332; PubMed=11872842;  
RA Gutman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,  
RA Greenberg J.T.;  
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant  
Pathogen Pseudomonas syringae";  
RL Science 295:1722-1726(2002).  
DR EMBL; AF458405; AAL84254.1; --  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 1160 AA; 126695 MW; C0B67EE28860B991 CRC64;  
Query Match 12.7%; Score 1198.5; DB 2; Length 1160;  
Best Local Similarity 28.8%; Pred. No. 5.6e-53;  
Matches 336; Conservative 225; Mismatches 476; Indels 128; Gaps 37;  
Qy 313 KGKLQAPD-----NPPALNTLLKQTLGKDTQHYLAHHASSDGSQHLLDNKGHLFD 364  
Db 58 QGTQLGADGKPDFSTFNTFCMAPLLSDILAKPGQTYLACQSQPDVQGHLLQSNRLLH 117  
Qy 365 IKSTATSYVLHNSHPGEIKKLAQAGTGSVDSKSGKISLGSQTQSH-----NKTM 417  
Db 118 LAODSSSLAVTRSS-----TATLAVEANTPPALDMQVDDHIDHTPDGRKS 164  
Qy 418 LSQGEARHSLTGIWQHPGAARP--QGESRLHDDKHILHLPGLVWQS---ADKQTHS 473  
Db 165 LELPKLAHLAHITGVHKNTPGDRLPVREERLRIHEDRLYQDPTGTFRWKAPHEMEIAFN 224  
Qy 474 QLSRQADGKLYALDNKRTLQNLSDNKSSEKLVDKISVSDQRGQVAILTDPGRHKMSI 533  
Db 225 SLATGNGSIYAQSD--VVVTLSPFPFHVKEVLTFSVAPDNTAALLS---GKETQAV 280  
Qy 534 M-----PSLDA--SPESHISLSLHFADAHQGLHKGSELEAQSVAISHGRLVVADEGKLF 587  
Db 281 LLDTMSPVIGSLTPKTKALELD-----HGLA--QAVAIGLSAKKLTVADNQGRLY 329  
Qy 588 SAAIPKQDGNELKMKAMPOHA---LDEHFGHDHQLSGFFHDDHGLNALVKNFRQOH 643  
Db 330 SAD--RSAFERNEPTRLMPERTHYTLAQAMGSHNSVTGFINGGDGRVHALIKNRQGEIH 388  
Qy 644 ACPLGND--HOPHPGNLTDALVIDNQLGLH--HTNPPEHILDMGHLSLALQEGKLYHFD 701  
Db 389 SHALDEQASKLESQNLNLTALVLDNTLGNMPTAPEPANRLNDRAGLVGSGGRIQWMD 448  
Qy 702 QLTGKWTGAE--SDCKQKKGLDGAAYLLKQGEVRLNI--NOSTSSIKHGTENVSFLPHV 758  
Db 449 ATPQCKWDAGIKDIDRLQRGADSNAYVLKGGKLLRLTVTPHPNQAFDHT--ALAQATR 506  
Qy 759 RNKEPFGDALQGLNKDKQAQAMAVIGVNYKYLALTEKGDIRSFOIKPTGQQLERAAQILSR 818  
Db 507 STRVEMKEIEGL--EDRVITAFAMVSDTRFVALDDKNLLT-----VHKQKHPPVPMDF 558  
Qy 819 EGISGLKLDHVDHKNLYALTHTHEGVFPHOPREAWQ---NGAESSSWHKLALPQSSKL 874  
Db 559 PGLGDIKELSLDEKHNLYARTSTGGFLCFLPKTAQWQANKLGLLEGKWTVPATQGO--PV 617  
Qy 875 KSLDMSHEHKPIATFEDGSQH---QLKAGGWHAAYAAPERGPLAVTSGSQTVFNRLQGV 931  
Db 618 KALFSDNDNLSAQIEDAPEQGLMQLKQGTWQTF---KQRP--VEENGLNDVHSKRKS 672  
Qy 932 XGKVIKPSGLTVKLSAQGTGKMGTAEGKVKSKSERIRAYAFNPTMTPTPRIKAAATQ 991  
Db 673 KTWRIPTGTLVTRDVMNIGLGGVKEKRNPSA--GELWRANVYKPDGKVPFRMKNIGNHIQ 731







Db 1126 ANFAPNVNVALGSLGSGGP--SVAGFALPFLSVVMVKFDRGESQAFSFGFKHPEPVTQSI 1183  
QY 1576 SELTSLGKHGKDSATTOMLAALKELDDAKPAEQHILQOHFSAKDVVG--DERYEAU-R 1632  
Db 1184 DDLKESVSAY---SSTFKHDLAGIDQSGTVDQQLATVQRFIAAHPMATKPDAYHAISQ 1240  
QY 1633 NLKULVIRQQAADSHMELGSAHSHSTYNNLRINNDGIVELLKHFDALPASSAKRLG 1692  
Db 1241 SLDKLMTQODLVKNGLRQLASVSSVTRVGL---RDDG---RHAWLDDVAPANKA-AIE 1292  
QY 1693 EMMNDPALKDIIKQLOSTPSSASVSMELKDLREOTEKAILDKGVGEEVGVLFQDRN 1752  
Db 1293 QWLKDDPQLAQVNLQOHGEGSVKLMELKQVLRITERRHLAGDNTLEPLRQALSDDR 1352  
QY 1753 NLRKVSVSQSVSKSEGFNTPA-LLLGTSNAAAMSMERNIGTINFKYQGOQONTERRFTL 1811  
Db 1353 NLRKVSMSLSYATQSHGMSIPAGANLSPSSAGLSYTKRVNADLEYGVDKFLRMNL 1412

RESULT 8  
Q888Y1 ID Q888Y1 PRELIMINARY; PRT; 1957 AA.  
AC Q888Y1;  
DT 01-JUN-2003 (TREMELrel. 24, Created)  
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Candidate type III effector HolPtor.  
GN HOLPTOR OR PSPT00883.  
OS Pseudomonas syringae (pv. tomato).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=323;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DC3000;  
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,  
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwin M.,  
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,  
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,  
RA White O., Fraser C., Collmer A.;  
RA "Complete sequence of Pseudomonas syringae";  
RL Submitted (MAR-2003) to the EMBL/GenBank/DDJB databases.  
DR EMBL; AE016858; AA054417.1; -.  
DR TIGR; PSPT00883; -.  
KW Complete proteome.  
SQ SEQUENCE 1957 AA; 210393 MW; 2B94903E5E07BAE3 CRC64;

Query Match 3.4%; Score 325; DB 16; Length 1957;  
Best Local Similarity 19.8%; Pred. No. 1.4e-07;  
Matches 388; Conservative 249; Mismatches 710; Indels 616; Gaps 90;

QY 208 IKPEPVGTSKATTAHADRVIAQEDDDSEFQQLHQRLARERENPPPKLGVATPISA 267  
Db 2 VKVTSSTGFANPLSHADSVSPA-----NSPPQLPE-----PV-- 34

QY 268 RFQKPLTAVAESVLEGTTDTOSPLQPSMLKSGAGVTPL-AVTLDKG---LQAPDNP 323  
Db 35 ----HLVDLSRSSKGGWRN---RPHASLSNQ---VLELQAVPSQKGVVRSHADGE 83

QY 324 PALNTLLKQTLGKTQHYLAHSHASDGSQHLLLNKNGHLFDIKSTATSYSLV-----H 376  
Db 84 SVINA-----WLAKRPSVQSETS--LNDGKGL--VRYTPVNHEPLAPRNEAFF 127

QY 377 NSHPG-----EIKKLAQAGTGSVSDGKGLSLSGTSQSHNKMTLSQPGEAHRSLLT 430  
Db 128 TSVFGMLMAVITVPEMEHGISGDITADAVARLIA-----EPPIGLLT 170

QY 431 GIWHPAGAAAPQGESIRLHDKIHLHPGLVWQSAADKTHSLSQASQADGKLYALKDNR 490  
Db 171 GIWQSSHDRAVLERGGV-VHTANMEERWAPLTPGINPREPLRMAGLQADGGVY-LNNGS 228

QY 491 TLQNLSDNKSSEKLVDKTKSYSVQORGOVALITDTPGRHKMSIMPSLDASPESHISLSH 550

Db 229 QLWRL-----TETRAESVTTENLPEGAAVRI-----GAGGEVH----- 261  
QY 551 FADAHQGLH-----GKSELEAQSVAI-----SHGRLVADSDGKFLPS 588  
Db 262 --GLHEGALHNGISIRPIELWRPKAGAFGRQSPARPVDLLPLPGGTAAIILDDKGRIVH 319  
QY 589 AAIKQGG--DGNELKMKAMPQHAIHDEHFGHDQIS--GFFHDDHGLQNALVKNFNQOCHA 644  
Db 320 ADLKTGTAVEAHLKLPKPA-----DFAQKGWAVTAMGLSRDDTVHL--MLQONQRRMSL 372  
QY 645 CPLGNDHGFHPCWNJTDALVIDNQGLGH-----HTNPEPHILDMGHGLGSLALQEG 695  
Db 373 QRAPGEALFRFAYLLDREPLLLYTEGLHVPSEAAVQSHVQLDGH--AOLGHI-----DG 424  
QY 696 KLVFDQLTKGW-----TGAE-----SCCKQ--LKKGLDQ-- 723  
Db 425 VLYHKAAPDQWERLKQSGGEPLTGLTALYSPLGFIIDRKPVFALVGARQVVELKBLGR 484  
QY 724 AAYLLKDGVEKR-----LNINQSTSSIKHGTENVFSLPHVRNKPPEPDALQGLNKDKA 777  
Db 485 TSWLPSDAELPRHPAGGFLAVIPDITVALRTSPIAQF-----DEPV 524  
QY 778 QAMAVIGVVKYALATEKDIRSQIKPTQTOQLERPAQTLRE-----GISBELKDIHVDH 832  
Db 525 QALAVHGNRRVVALTDSGLMAADADTPARL-----PTLQRPPIATAVGLNDQLLVHHPH 580  
QY 833 KQ-----NLVALTHEGEVFHQPREAWNGAESSWH-KL 865  
Db 581 SQRQLKLSAKDDWEPVPIILGLVHPSSIRATRTGQIQVQLGENWHITLPSMTSHDNQ 640  
QY 866 ALPOSESKLSLSDMSHEHKPIATFEDGSOHLKAGWHAYAAAPERGLAVGTSGSQTVFN 925  
Db 641 RLP---ARVKP-EPEGDSEANFLAGSNALANQQAASRISTPHHDA SVVITLACTANN 696  
QY 926 RLMQGVKGVIPGSLTVKLSAQ-----TGMWTAEGRKV---SSKFSERIRAVAFNPTM 977  
Db 697 -----PLTWASSLOAVDVTTRAQVGALARDVVGAANSTMRMAHTLGVVL 742  
QY 978 -STPRPIKNAAY---ATQHGQGR---EGLKPLVEMOGALI---KQLDAHVNRNAPQPD 1027  
Db 743 PPTQEKRLASFHNEAKQAYTSGKILFEHLPSLAQVRVASAVGSDGSRFGLSHOQTQ-- 800  
QY 1028 LQSKLETLDLGEHGAELINDMKRFRDELEQSATRSVTVLGQHGVY-----KSNGEI 1079  
Db 801 --RLTLTLREGKLEA-LLRDLRK-----IGFHEGVIMGDMGSDSAHGLV 841  
QY 1080 NSEFKPS-----PGKALVGSFNVNRSGDLSKLSQAQVHAHTPPSA----- 1119  
Db 842 STTSTPTFLAELWRRQHSRVDKAL--SSAGLSRS-EDIFPDNLINLINALAGGAALNDRM 899  
QY 1120 --ESKLQSMI---GHFVSAGVDMSHQKGEIPLGRQRPNDKTKALTKSRLILDVTTIGE 1172  
Db 900 SERAEALLSVLCEVSEKMMRAGVRLPADDGVSVDSAHSQAP-----YGLRTAGL 947  
QY 1173 LHELADKAKLVSDHKPDADQIKQLRQOFDTLREKRYESNPVKHYTDMGFTNKALEANYD 1232  
Db 948 IAGLVVDYDALLS--STDAQALE-----MAERLQDARLAALCKLGLSSWGQAAFPDD 997  
QY 1233 AVKAFINAFKEHGVNLTTRTVLESQGSABLAALKKNTLLSLDSGSMFSRSYGGGVS 1292  
Db 998 VVITTF-----REQISLPGSARRTQLKNLGLPDPDAAPDEMAARM----- 1036  
QY 1293 TVFVPTLSKKVPVPIPGAGITLDRAVNL--FSRTSGGLNV--SFGK-----DGG 1339  
Db 1037 -----SOLLLD-LFNRTFTTQSGRLGSLGSADMKHLNAPSVG 1077  
QY 1340 VSGNIMVATGHDVMPYMTGKKTSA-----NASDMLSAKHKISPOLRIGAAVSGTQGG 1392  
Db 1078 VTGALQVLG--VERIGDKGKDAGLVAFFVRHAKASVSATSGIGIDFKPGPGTGGRV-- 1133  
QY 1393 TLQNSLKFKLTDELPQFIHGLTHGTLTPAELLQKQIE--HMQKGS----- 1437  
Db 1134 -----IDSRPGRSMNNTWGGST-----NLGISGAYOHGQGAIVLAPSTISDFV 1177



Db 871 LRGAIDRMWG-----ALAQIGVAAARVLGPDLTLDAAAGRRKQRRPAKADDP 918  
Qy 1119 AE--SKLOSMGLGHFVSAGVMSHOKGEIPLGRQDPNDKLTALTKSRLSLDVTVTIGELHEL 1176  
Db 919 DDLPLKMSWIDFV-----RAE-----AQKPSPEAGAKAPAPVDFPARLGDHRI 966  
Qy 1177 ADKAKLVSDH-----KPDADQIKOLRQOF----- 1200  
Db 967 VELMHTLSEHGKVPAPDLTGQDTRDRFAILTGAVGRHLLTLDLATRLVLTLPQTVAVIG 1026  
Qy 1201 -----DTRERKYESNPVKHYTDMGTGFKANKLEANYDAVKAFINAPKKEHGVN---ITT 1252  
Db 1027 SDSAVEVLRSQ--VETDKVLRILAREFGSNWHDAAEFWETAQTFKKEVWMPKPSFNRKFAES 1085  
Qy 1253 RTVLSEQSAELAKKLKNTLLSLDSGESMSFSRSYGGGVSTVFVPTLSKKVPVPVPCAG 1312  
Db 1086 HAVRDASSPAEMALKFAEMAKGLNRSSTLFIEKSGASGVTAGVLPKRVGAGFAV 1145  
Qy 1313 ITLDRAYNLSFRTS-----GGLNVSGRGGVSGNIMVATGHDVMTYMTGKTSAGNAS 1367  
Db 1146 LGVDRTTWGVVERTADKLAEGPLVAFFVRQSN-----KGLTWGGGGM 1188  
Qy 1368 DWLSAKHKISDRLRIGAAVSTLQGTLONSLKFKLTEDELPCFIHGL-----THGTLTPAE 1423  
Db 1189 DFKPLRKVLGFRLOGGAQASAVMH--GKGAAMLVAPDNIDEFARRLPDPASH---DPCR 1243  
Qy 1424 LLOKGIHOMKGGSKLTFSVDTSANLDRAG-----INLNEGSKPENGVT 1469  
Db 1244 LLELG-----INQGAIGLDMIEQQNLNSANAGLGGYADKIPGFPAHRHGGDS---STSA 1296  
Qy 1470 RVSAGLS---ASANLAAGSRE-----RSTTSQFGSTTSASNRRPTFLNGVGAGANL 1518  
Db 1297 FQGTGLQGFVGANVMWGVDFELKLQHAWEPIGCV-----EYQGGRWSSANA 1344  
Qy 1519 TAALGVAHSTHEGKPVGIFP-----AFT-----STNVS---AALALDNRTSOSISLELKRA 1567  
Db 1345 FASL-----VQOG---GLLPHVSDAFTLVLRSLNITLLGASVELSGVESFKRTLDMKTA 1395  
Qy 1568 EPVTSNDISELTSTL-----GKHFQDSATTKMLAALKELDDAKPAQLHLIQ 1614  
Db 1396 ARVTPQWESQAALAREVFPQATGHGDPHLKAIATLQAA-----K 1439  
Qy 1615 QHFSKADVGDEREAVENLKKLVIRQO--AADSHSMELGSAHSTTYNNLSRINNDGIVE 1673  
Db 1440 STWAAR--TEHERASFVDRAEQLLLDQDLASGRAMLPGAKIERNIPFRSLVDTRKNS 1497  
Qy 1674 LLKHFDALPASSAKRLGEMMNNDPALKDIIKQLQSTPPSSASVSMELKDLRE----- 1728  
Db 1498 KAHRSGLALMEA--AERAREAV--PGLADAMR-----AMSERDGVNDVRFVP 1540  
Qy 1729 -----QTEKAILDGKVGREEVGLFQD--RNLRVKSVSVSQSVKSEGN-TPALLL 1778  
Db 1541 QMPSYINAVNRLMEGLKSWAEFNTWARTVPAPYRLTEICAKDSNRSFTLNPLPL 1600  
Qy 1779 GTSNSAAMSERNITGTFNKYG 1800  
Db 1601 AFNDSAEVSRSLFAEVHLRYG 1622

## RESULT 10

Q8NWQ6 PRELIMINARY; PRT: 9904 AA.  
AC Q8NWQ6;  
DT 01-OCT-2002 (Tremblrel. 22, Created)  
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)  
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)  
DE Ehb protein.  
OS Ehb OR MW1324.  
GN Staphylococcus aureus (strain MW2).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI\_taxID=196620;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=22040717; PubMed=12044378;  
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,  
RA Nagai Y., Iwama N., Asano K., Naimi T., Kuroda H., Cui L.,  
RA Yamamoto K., Hiramatsu K.;  
RT "Genome and virulence determinants of high virulence community-  
RT acquired MRSA.";  
RL Lancet 359:1819-1827(2002).  
DR EMU; AP004826; BAB95189.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR002988; GA.  
DR InterPro; IPR005977; Gpos\_Ysirk.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF01468; GA; 48.  
DR Pfam; PF04650; Ysirk signal; 1.  
DR TIGRFAMs; TIGR01643; YD repeat\_2x; 8.  
DR TIGRFAMs; TIGR01168; Ysirk signal; 1.  
KW Complete proteome.  
SQ SEQUENCE 9904 AA; 1068476 MW; CE86908F75EA56B2 CRC64;

Query Match 3.0%; Score 283.5; DB 16; Length 9904;  
Best Local Similarity 18.6%; Pred. No. 0.0003;  
Matches 396; Conservative 306; Mismatches 817; Indels 605; Gaps 94;  
Qy 8 TEHKAAYHTAAHNPVGHGVALQQ--GSSSSSPQNAASLAAGKMRGMKPRIHQHPSTAAD 65  
Db 2543 TQYNAIHNAQQQINTAKTEAQVINDRATPQQVSDALT-----KVRAAQTKID 2592  
Qy 66 GISAAHQKGSFSLRGCLGTTKFKRSAPQGGPQGTTHSGATRLDLLARDGGETHEAAP 125  
Db 2593 QAKALLQNKEDNSQ---LVTSKNNLQSSVNVQVST--AGMTQQSIDNNAKKREATEIT 2647  
Qy 126 DAARLTSSGGV-----KRRNMDDMAGRPMVKGSGEDKVPQOKRHQLNNFQMQQT 177  
Db 2648 AQRVINDGATPQOISEKRVNALTALNQAQKNTADTHTLEQAVQQLNRTG---T 2703  
Qy 178 MLSKMAHPASA---NAGDRLQHSPPHIPGSHHEIKEEFVSGSTSKATTA--HADRV--- 227  
Db 2704 TTGK--KPASITAYNNSMHALQASLTSAKNSANALIQKPIRSVQEVQIATNVRNREL 2761  
Qy 228 -----ETAEEDDDSEFQOLHOORLARENPQPKLGAVATPI SARFQPKLTAV--- 276  
Db 2762 TQAINQLVPLADNSAL-----RTAKTKLD---EINKSVTTDGMTQSSIQAVENAKRA 2811  
Qy 277 -----AESVLEGTDTTQSPKQPQ-----SMLKSGAGVTPPLATLDKGLQLAPD 321  
Db 2812 QGTSTNAQVNNINGDATDQOIAEBKVEKYNLQKQIAGLTPLDPLAPLOFTAKTQLOND 2871  
Qy 322 -----NPPALNTLLK--QTLGKDTQHYLAH-----ASSD 349  
Db 2872 IDQPTSTTGMTSTSIAPNEKLSAARTKIQIEDRVLASHPDVATIRQNVTAANAATAKD 2931  
Qy 350 GSQHLLDNKGHLDPDKS-----TATSVLHNSHPGETKGLAQAGTGSVSDGKSG 402  
Db 2932 QARNGLTVDKAPLENAKQLQHSIDTQSTTGMTQDSINAYNAKLTAARNKIQIIN----- 2987  
Qy 403 KISGSGTQSHNKTMSPGCEAHSLLTGIWHQFAGARPQGESIRLHDDKHILHPLG 462  
Db 2988 QVLAGSPTEVQINTNTSAANOAKSDL-----DHARQALTP-----DKAPLOTAKTQ 3033  
Qy 463 VWSADKDT-----HSQLSRQADGKL-----YALKDNRTLQNLSD-----NKSS 501  
Db 3034 LEQSINQPTDTGTMTASLNAYNQKLAARQKLTEINQVLNGNFTVQINDKVTENQAK 3093  
Qy 502 EKLVDKIKSVSDQGVQVAILTDPGRHKMSIMPSLDASPEHSLSLHFADAFHGLHG 561  
Db 3094 DQLTARQGLTDRQPALTTL-----HGAS---NLQAQONNFTQOINAAQHAALETI 3144  
Qy 562 KSELEAQSVASHGRVLVADSEGLFSAAIPTKQDGNELKMKAMPQALDEHFHGHQIS 621  
Db 3145 KSNITALTAMTKLQDSVAD-----  
Qy 622 GFFHDDHGQNLALVKNNFRQOHACPLGNDHQFHPGWNLTDALVIDNQGLHHTNPPEHI 681

```
Db 3165 -----NNTIKSQNVTDATP-ANKQAYDNVNAAGVIGET-----TNP-----T 3203
Qy 682 LDMCHGLSLALQBEKHLHYFDLTGKTGABSDCKQLKKGLDGAAYLLKQEVKRLINQS 741
Db 3204 MDVNTYNOKA-----ASYKSTKDALDGOONLORAK-----TEA 3236
Qy 742 TSSIKHGTENVFSLPHVRNKPEPDALQGLNKDDKAQAMAVIGVKNKYALTEKGDIRSFO 801
Db 3237 TNAITHASD-----LNAQKNALTOQVN-----SAQNVHAVN-----D 3269
Qy 802 IKPGTQQLERPAQTLREGISGBELKDHYD-----HKONLY--ALTHEGEVPHOPREA 852
Db 3270 IKQTTQSL-NTAMTGLKRGVANHNVQVSDNYVYVADTNKNDYNNAYNHANDII----- 3322
Qy 853 WQNGARSSWHKIALPQSESKLSLDSHEHKEPIATFEDGSQHLKAGGWHAAAPERGP 912
Db 3323 --NG--NAQHPVITPDSVNNALSNVTSKEH-----ALNGEAKLNAKQAEAN 3364
Qy 913 LAVTSGSQTVFNRLMQGVKGVIPGSL-----TVK-----LSAQTTGGMTCABGRK----- 959
Db 3365 TALGQLNN--LNNARQNLQSQINSAHQIETVNTIKQATNLNSAMGNLRQAVADKQOVK 3422
Qy 960 -----VSSKFSERIRAYAFNPTMSTPRPTKNAAYATQHGQGBEGL 1000
Db 3423 RTEDYADADTAKQAYNSAVSS--AETIIINQTTNPTMSVD--DVNRATSAVTSNKNALNGD 3479
Qy 1001 KPLYEMOGALIKOLDHNVHNAPOPLQSKLETLD-----LGEHGAEL----- 1044
Db 3480 EKLAQSKTDARAIDALPHLNNQAKADVKSINAASNIAGVNTVKQGTDLNTAMGNLQ 3539
Qy 1045 -LND-----MKFRDELEQSATRSVTVLGHQGVL--KSNGEINSEFPKSPGKALVQSFN 1096
Db 3540 AINDEQTLINSQYQDATPEKKTAYTNAVQAADILNKSNGQ--NKTQDKQVTEAMNQ-LN 3596
Qy 1097 VNRSGDLSKSLQOAVHATPPSAESKLSQMLGHFVSAGVMSHOKGEIPLGRORDPDNKT 1156
Db 3597 SAKNLDGTRLLDQA-----KOTAKQQLNNMT-HLTTA-----QKT 3631
Qy 1157 ALTKSRLILDTVITGELHELADRAKLVSDHKPDADQIKQLRQOFTDLTREKYESNPVKHY 1216
Db 3632 NLTNQ--INSGTTVAGVHTVQSN-----NTLDQAMNTLRQSIANKDAPKAS 3676
Qy 1217 TDMGFTHNKALEANYDAVKAFFINAFKKEHGVNLTTRTVLESQGSASLAKKNTLLSLD 1276
Db 3677 EDY-----VDANDKQYANNVAAAEIINANSNPENPSTITQKAEQVNSKTKALN 3729
Qy 1277 SGSMSPSRYSYGGVSTVFVPTLSKKVPVPIPGAGITLDRAYNLSPRTSGGLNVSPGR 1336
Db 3730 GDENLATAKQN-----AKTYLNTLT-----SIT-DAQKNLISQISSATFVS-GV 3772
Qy 1337 D-----GGVSGNIMVATGHDVMPYMTGKTSAGN-----ASDWLSAKHK 1375
Db 3773 DTWKQNAQHLDDQAMSLQSGINNESQVKSSEKYEDADTNKQOQYDYNAITAKALNKKQH- 3831
Qy 1376 ISPDL-----RICAAYS--GTLQGTQLNSLKFELTDELPGFTIHTGLTTPAEILLQKIE 1430
Db 3832 -GPNTAQNAVEAALQRYNTAKDALNGDAKLIQAQNAKQHLGTLTH--ITTAQ--RNDLT 3886
Qy 1431 HQMKQSGKL--TFSVDTSAN-LD-----LRAGIN-----LNEGSKPQNGVTVARV 1471
Db 3887 NQISQATNLAVESVQKNANSGLDAMGNLQTAINDKSGTTLASQNFLLDADEQKNAYNQAV 3946
Qy 1472 SAGLSASANLAAGSRERSTTSQFGSTTSASNNRPTLVG-----GAGANLTAALGVAH 1526
Db 3947 SAA--ETIILKQGTGNTAKTAVEQ-----ALNNVNNAKHALNGTQNLNNAKQAAITANGASD 4002
Qy 1527 SSTHEGKPVGIPPAFTSNVSAALALDN-RTSQSISLELKEAPEVTSNDISELSTL-GK 1584
Db 4003 LNOHKQDAL-----KAAQANGAQRVSAQDVQRNATELNTAMGTLKHAIADKNTLASS 4055
Qy 1585 HFKDSATTKMLAALKELDDAK-----PAEQHLILQOHFSAK-DVVGDER----- 1627
```

```
Db 4056 KYVNADSTKQNAVTTKVTNAEHIISGTPVTVPVTTPTSEVTTAAANQVNSAKQELNGDERLVA 4115
Qy 1628 -----YBAVRNL-----KKLVIRQQAADSHSMELGSAHSHTTYNNLSRINNDGIVELLHK 1677
Db 4116 KQNAVTAIDALTQLTNPQAKLKEQVQANTLD-----DAMNSLOQALNDKDATLRNQ 4168
Qy 1678 HFDALPASPASAKR-----LGEMMNDPALKDIIKQLOSTPFPSSA----- 1716
Db 4169 NY---LDADESKENAVTQAVTAAGILNKQTCGNTSKADVNDALNAVTRAKAALNGAENL 4225
Qy 1717 -----SVSMELKDLGREOTEKA-----ILD-GKVGREEVGVL---FQ 1749
Db 4226 RNAKTSATNTINGLPLNLTLQKDNLKHQVEQANVAGVGVKDKGNTLNTAMGALRTSIQ 4285
Qy 1750 DRNLRKVSQSVQSVKSEGFNTP-----ALLIGTSNAAAMGWERNIGTFNKYGGDQN 1804
Db 4286 NDMTTKTSQNYLDASDSKNKNYNTAVNNANGVINATNPN---NWDAN--AINGMANQVNT 4340
Qy 1805 TPRRFTLEGGIAQANPOVASALTD 1828
Db 4341 TKAALNGVQNLQAQAKTNAATNTINN 4364
```

## RESULT 11

```
Q9LAW1 ID Q9LAW1 PRELIMINARY; PRT; 57 AA.
AC Q9LAW1;
DT 01-OCT-2000 (TremBrel. 15, Created)
DT 01-OCT-2000 (TremBrel. 15, Last sequence update)
DT 01-OCT-2000 (TremBrel. 15, Last annotation update)
DE Hrp-secreted pathogenicity/avirulence protein DsPE (Fragment).
GN DsPE.
OS Erwinia amylovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Erwinia.
OX NCBI_TaxID=552;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=Ea246;
RA Kim J.F., Laby R.J., Beer S.V.;
RT "Comparison of the hrpN-flanking regions of two Erwinia amylovora
RL strains with different host specificity.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RD EMBL; AF083620; AAF63403.1; -.
FT NON TER 57
SQ SEQUENCE 57 AA; 5788 MW; AFFOCA36311E4BE1 CRC64;
```

Query Match 3.0%; Score 283; DB 2; Length 57;  
Best Local Similarity 98.2%; Pred. No. 5.8e-08;  
Matches 56; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MELKSLGTEHKAAVHTAAHNPVGHVVALQGGSSSSSPQNAASLAABGKNGKMPRI 57

Db 1 MELKSLGTEHKAAVHTAAHNPVGHVVALQGGSSSSSPQNAASLAABGKNGKMPRI 57

## RESULT 12

```
Q99U54 ID Q99U54 PRELIMINARY; PRT; 6713 AA.
AC Q99U54;
DT 01-JUN-2001 (TremBrel. 17, Created)
DT 01-JUN-2001 (TremBrel. 17, Last sequence update)
DT 01-JUN-2003 (TremBrel. 24, Last annotation update)
DE Eha protein.
GN Eha OR SA1267.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N. A.
RA MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
```





QY 979 TPRPKNAYATQHWQREGKPLIYEMQCALIKOI-----DAHVNRHAP 1024  
Db 854 ----HNGAGATEAGKRLGVYDEKHLNTNEHLKTQLVETGREHVIDYFAFGHEHLLREGT 909  
QY 1025 OPDL-----QSKLETLDLGEH-----GAEILLDMK 1049  
Db 910 QHELGSVYNDESRLRTPDGAHWHYDYKVTQKTQVTTAPAKIISGNDITIDGK 969  
QY 1050 R-FRDELEQSTRSVTVLQHQG-----VLKSGNEINSFEPKPS-PGKALVQSFN 1096  
Db 970 EVENTDSQIIAGGLIVQTEKDLHNEQTFGEKKVSENGKLHYSYREKHGR---DSFG 1026  
QY 1097 VNRSCDLSKSLQAVHATPSAESKLSMLGHFVSAGVDMHQGEIPLGRQRPDKT 1156  
Db 1027 HSEQNYTLPEBITRINISLGSFAYESHKALSHAPSQGTLPQSGN---ISLPTYSNFT 1083  
QY 1157 ALTKSRL-ILDTVTIGELHELADKAK-----LVSPHKPDA---DQIKQLRQQFDLREKR 1207  
Db 1084 PLPSSSLIINPVNGYLVETDPRPANYRQWLGSDYMLDSLKLDPNHLHRLGDDGYEQR 1143  
QY 1208 YESNPVKHYTDMGFTHNKALEANYDAVKAFINAFKEHGHVNLTRTVLESQGSALAKK 1267  
Db 1144 LINEQIAELT--GHRRLDGYQNDDEQFKALMDNGATAARSNNLSVGIAL----SAEQVAQ 1197  
QY 1268 LKNTLLSLDSGESMSFSRSYGGGVSTVPPLSLKKVPVPIPGAGITLDRAVNLFSRST 1327  
Db 1198 LTSIDLVLVQKE-----VKLPDGGTQVLVPQYVRVKNGDIDKGAL-----LGSNTQ 1247  
QY 1328 GGLNVSGFRDGGVSGNIMVATGHDVMPYMTGK---KTSK-----GNASDWLSAKHKIS 1377  
Db 1248 INVSGSKNSGTIAGRNALIINTDLNIGRIHQAOKSAVATQDINNIGMLSAEQTLL 1307  
QY 1378 PDLRIGAAV-SGTQGLTQNSLKFKLTEDELPG-FIHLGTHGLT-----PAELLKGIE 1430  
Db 1308 --LNAGNINQSSTASSQNTQSGSTYLDRMAGIVITCKEKGVLAAQAGKDIINIAGQIS 1365  
QY 1431 HOMQGSKLTVSDTSANL-----RAGINLEDGSPNGVTVARVSAGLSASA----- 1479  
Db 1366 NQSBQG-QTRLOAGRDINLDTVQTSKHOATHFDADNHVIRGSTNEVGSSIQTGDDVTLLS 1424  
QY 1480 --NLAAGSRERSTTSGFGSTTSASNNRPTFLNGVAGANLTAALGVAHSSTHEGKPVGI 1537  
Db 1425 GNNLNKAAAEVSSANGTL--AVSKND-----IN-ISAGINTT---HVDDASKHTGRSGG- 1473  
QY 1538 FPAFTSTNVSAALADNRTSISLELKRABPVTNSIDELSTL-----GKHFK 1587  
Db 1474 ---GNKLIVITDKASHHETAQSTFEGKVVLQAGNDANILGNSVSDNGTQIQAGNHVR 1530  
QY 1588 DSATTKMLAALKELDDAKPAQLHLQHFSAKDVGDYERVENLKKUVIROQAADSH 1647  
Db 1531 IGT-----QTSQSETHYHQKSGLSAGIG-----FTI---GSKTN 1565  
QY 1648 SMELGSASHSTYNNLSRINNDGIVELLKHFDALPASSAKRLGEMM-----NNDPALK 1702  
Db 1566 TQENQSQSNEHTGTVGSLKGDITI-VAGKHYE-----QIGSTVSSPEGNTIYAAQ 1615  
QY 1703 DIIKQLOSTPSSASVSMELKDXLREQTEKAILDKGVGREVG-----LFQDRNNLRV 1756  
Db 1616 SIDIOAAHNLNSNTQTYEQKGLTVAFSSPVD--LAQAIIVAQSSKQVQSKND-RV 1672  
QY 1757 KSVSVSQS--VSKSEGNTPALLIGTSNAMSERNIGTTFNFKYQDQNTPRFT-LEG 1813  
Db 1673 NMAAANAGWQAYQTGSAQNLANGTNGTNAKQVS-----ISITYGEQON--RQTTQVQA 1723  
QY 1814 GIAQNPQVASALTDLKEGLEMSK 1838  
Db 1724 NQAASQIQAGGKTTLIATGAEQS 1748

RESULT 14

Q931R6

ID Q931R6

AC Q931R6;

PRELIMINARY; PRT; 6713 AA.

DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE Hypothetical protein ebhA.  
DE EBHA OR SAV1434.  
GN Staphylococcus aureus (strain Mu50 / ATCC 700699).  
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
OX NCBI TaxID=158878;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21311952; PubMed=11418146;  
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,  
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,  
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,  
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,  
RA Sekimizu K., Hirakawa H., Kubara S., Goto S., Yabuzaki J.,  
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,  
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,  
RT "Whole genome sequencing of methicillin-resistant Staphylococcus  
aureus.";  
RL Lancet 357:1225-1240 (2001).  
DR EMBL; AP003362; BAB57596.1; --  
DR InterPro; IPR002988; GA.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF01468; GA; 46.  
DR TIGRFAMs; TIGR01643; YD repeat 2x; 7.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 6713 AA; 722306 MW; BBCC0536AC341BF5 CRC64;

Query Match 2.9%; Score 273.5; DB 16; Length 6713;  
Best Local Similarity 17.8%; Pred. No. 0.00051;  
Matches 407; Conservative 322; Mismatches 904; Indels 651; Gaps 95;

QY 11 KAAVHTAAHNPVGHGVAL-----QGGSSSSPQNAASLAAGKNGKMPRIHQPSTA 63  
Db 556 KAAVENALSQVTNAKAGLNGNHNLEQAKSNANTINGLOHLLTAKOKLKQ-QVQQAQNV 614  
QY 64 ADGISAHHQKKSFLRGLGTGKFSRPAQGGTTHSKGATLRLDLARDDETQHEAA 123  
Db 615 A-GVDTV--KSSANTLNGAMGTLSNI-----QNTATKNGQNYLD--ATERKNYNNNA 664  
QY 124 APDAARLTRSGVKKRNMMDMAGRPVKGSGEDKVPQT--QKRHL---NNFGQMRTM 178  
Db 665 VDSANGVINA--TSNPNMD-----ANAIQIATQVTSTKNALDGTENLTAQAKQTA 712  
QY 179 LSKMAHPASANAGDRLQHSPPHIGSHHEIKEEPVGSTSKATTAAHADV-----ETAQ 231  
Db 713 TNAIDGATNLNKAOK-----DALKAQVTSQORVANVTSIQQTAN 751  
QY 232 E-----DDSEFQQLHQRLARENPFPKLGVAATPISARFQPKLT--AVA 277  
Db 752 ELNTAMQLQHGIDENATKQTKYRDAEQSKKTAQDAVAAAKAILNKQTSNDRKAAV 811  
QY 278 ESVLEGTTTQSPKLPKPSMLKSGAGVTPIAVLTDKGLQLAPNPPALNTLLKQTLGKD 337  
Db 812 DRALQQTSTKDALNGDAKLAEAAXAARQNLGTLN---HITNAQRTALEQIQIQAATVD 867  
QY 338 TQVLAHHASS-DGSQHLL-----LDNKGHLFDIKSTATSYSVLHNSHPEIGKG 386  
Db 868 GVNTVKTNTALDGMNSLQCAINDKATLRNQNYLDADESKRNWAYTQAVTAABEG-ILNK 926  
QY 387 LAQAGTGSVSDGSGKISLGSQTSHNKTMLSQGEAHRSL--LTGIWHPAGAAARPOG 444  
Db 927 QTGNTSKADVNDALNAVTRAKAALNCAENLRNAKTSATNTINGLPLNTLOKQNLKHQV 986  
QY 445 E-----SIRLHDDKIHLHPGLVWQS-----ADKTHSLSQSRQA 479  
Db 987 EQAQNVTGNGVGVKDGKNTLNTAMGALRTSIQNDNTTKTSQNYLDASDSKNKNYNTAVNNA 1046  
QY 480 DGKLYALKD-----NRT-----LQNLSDNKSSEKLVDKI-KSVSVQORG 517  
Db 1047 NGVINATNPNPDANAINDMANOVNTTKAALNGAQNLAQAKTN--ATNTINNAQDLNQKQ 1104





[illegible]

Search completed: July 7, 2004, 15:20:13  
Job time : 95.8129 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:06:21 ; Search time 7.80425 Seconds  
(without alignments)  
5032.400 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLOYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	AAW98012	Hypersens
2	715	100.0	139	AAy71096	Erwinia a
3	715	100.0	139	AAy84857	A hyperse
4	715	100.0	139	AAU02879	Erwinia a
5	715	100.0	139	AAE16450	Erwinia a
6	82	11.5	792	ABU20768	Protein e
7	81.5	11.4	402	ABU40422	Protein e
8	80.5	11.3	368	AAW89180	Amino aci
9	80.5	11.3	766	AAU07739	A snake v
10	80.5	11.3	787	AAU07740	A snake v
11	80.5	11.3	820	AAU07741	A snake v
12	79.5	11.1	957	AAU78250	Endometri
13	79	11.0	789	AAU81193	Mycobacte
14	76.5	10.7	1105	AAW44864	Human TPC
15	76.5	10.7	1105	AAW73958	Human TPC
16	76.5	10.7	1105	AAE15161	Human TPC
17	76	10.6	1235	ABG61674	Cadherin-
18	75.5	10.6	966	ABG93047	S. cerevi
19	75	10.5	243	AAW38542	Streptoco
20	75	10.5	317	AAU85961	S. pneumo
21	75	10.5	317	AAU13516	Streptoco
22	75	10.5	317	AAU60873	Mevalonat
23	75	10.5	317	ABU00733	S. pneumo
24	75	10.5	317	ABU97196	Enzyme po
25	75	10.5	2053	ABB71118	Drosophil

26	74.5	10.4	1471	4	ABB58739	Abb58739 Drosophil
27	74	10.3	422	6	ABU41136	Protein e
28	73.5	10.3	984	7	ADE38441	Human pro
29	73.5	10.3	2032	6	ABR43630	Mouse CLA
30	73	10.2	560	5	ABP39498	Staphyloc
31	72.5	10.1	421	6	ABU41975	Protein e
32	72.5	10.1	466	4	ABG63634	Human gas
33	72.5	10.1	590	4	ABG05720	Novel hum
34	72	10.1	584	4	AAU82869	S. epider
35	71.5	10.0	511	7	ADB64185	Human pro
36	71.5	10.0	920	2	AAW82500	Human OGT
37	71.5	10.0	920	2	ADE80755	Microsate
38	71.5	10.0	920	8	ADE76911	Human pro
39	71.5	10.0	967	6	ABR41060	Human MAP
40	71.5	10.0	1036	6	ABG71215	rat O-lin
41	71.5	10.0	1036	7	ADE62982	Rat Prote
42	71.5	10.0	1591	6	ABR41061	Human MAP
43	70.5	9.9	103	7	ADC95750	E. faeciu
44	70.5	9.9	670	4	AAU04025	Human pro
45	70.5	9.9	988	5	ABG61693	Cadherin-

## ALIGNMENTS

RESULT 1  
AAW98012  
ID AAW98012 standard; protein; 139 AA.  
XX  
AC AAW98012;  
XX  
DT 21-JUN-1999 (first entry)  
XX  
DE Hypersensitive response elicitor DspF.  
XX  
XX Hypersensitive response elicitor; DspF; disease resistance;  
KW Insect resistance; biological control; transgenic plant.  
XX  
OS Erwinia amylovora.  
XX  
PN WO9907206-A1.  
XX  
PD 18-FEB-1999.  
XX  
PF 24-JUL-1998; 98WO-US015426.  
XX  
PR 06-AUG-1997; 97US-0055105P.  
XX (CORR ) CORNELL RES FOUND INC.  
XX Bogdanove AJ, Kim JF, Wei Z, Beer SV;  
XX  
XX WPI; 1999-180362/15.  
XX N-PSDB; AAX24811.  
XX  
XX Nucleic acid encoding hypersensitive response-eliciting protein - used to improve growth of plants and impart resistance to disease and insects.  
XX  
XX Claim 18; Page 57-58; 75pp; English.

This polypeptide comprises the 16 kDa hypersensitive response elicitor protein DspF of Erwinia amylovora. The nucleotide sequence of the dsp region of E. amylovora strain Ea321 was determined using subclones of pCPR430. A two-gene operon was discovered comprising dspE (see AAX24810) and dspF (see AAX24811). The isolated dsp DNA molecules and encoded proteins can be used to impart disease resistance to plants, to enhance plant growth, and/or to control insects on plants. This is achieved by applying a hypersensitive response elicitor protein or polypeptide in a non-infectious form to plants or plant seeds, or by producing transgenic plants or plant seeds transformed with DNA encoding a hypersensitive response elicitor. Protection can be provided against a wide range of viruses, bacteria, fungi and insects, e.g. tobacco mosaic virus and tomato mosaic virus, Pseudomonas syringae, Xanthomonas campestris,

CC Fusarium oxysporum, Phytophthora infestans, armyworm, diamondback moth,  
CC etc. The method avoids use of infectious agents or polluting chemicals.  
CC Claimed transgenic plants are infected from alfalfa, rice, wheat, barley,  
CC rye, cotton, sunflower, peanut, corn, potato, bean, pea, chicory,  
CC lettuce, endive, cabbage, brussels sprout, sweet potato, beet, parsnip,  
CC turnip, cauliflower, broccoli, turnip, radish, spinach, onion, garlic,  
CC eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber,  
CC apple, pear, melon, citrus, strawberry, grape, raspberry, pineapple,  
CC soybean, tobacco, tomato, sorghum, sugarcane, Arabidopsis thaliana,  
CC Sainpaulia, petunia, pelargonium, poinsettia, chrysanthemum, carnation  
CC and zinnia  
XX  
SQ Sequence 139 AA;

Query Match 100.0%; Score 715; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 4.2e-78;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRILE 60  
Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRILE 60

QY 61 ADPQTSTILYSMLLQNFEMAAARGCWALDELHNRLCFQOSLEHLDEASFSDIYSGFI 120  
Db 61 ADPQTSTILYSMLLQNFEMAAARGCWALDELHNRLCFQOSLEHLDEASFSDIYSGFI 120

QY 121 EHAAEVREYIAQLDESSAA 139  
Db 121 EHAAEVREYIAQLDESSAA 139

## RESULT 2

AAV71096  
ID AAY71096 standard; protein; 139 AA.  
XX  
AC AAY71096;  
DT 08-SEP-2000 (first entry)  
XX  
DE Erwinia amylovora hypersensitive response elicitor encoded by dspf gene.  
XX  
KW Hypersensitive response elicitor; environmental stress resistance; plant;  
KW dspf gene.  
XX  
OS Erwinia amylovora.  
XX  
PN WC200028055-A2.  
XX  
PD 18-MAY-2000.  
XX  
XX 04-NOV-1999; 99WO-US026039.  
XX  
XX 05-NOV-1998; 98US-0107243P.  
XX  
PA (BDEN-) EDEN BIOSCIENCE CORP.  
XX  
XX Wei Z, Schading RL;  
XX  
DR WPI; 2000-376566/32.  
DR N-PSDB; AAD00671.  
XX  
XX

PT Application of a hypersensitive response elicitor protein to plants to  
PT impart stress resistance.  
XX  
XX Disclosure; Page 21; 84pp; English.

XX The patent discloses a method to impart stress resistance to plants by  
CC applying a hypersensitive response elicitor in a non-infectious form to a  
CC plant or seed. The present sequence is a hypersensitive response elicitor  
CC encoded by dspf gene from Erwinia amylovora. The protein is used to  
CC impart stress resistance to plants  
XX  
XX Sequence 139 AA;

Query Match 100.0%; Score 715; DB 3; Length 139;  
Best Local Similarity 100.0%; Pred. No. 4.2e-78;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRILE 60  
Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRILE 60

QY 61 ADPQTSTILYSMLLQNFEMAAARGCWALDELHNRLCFQOSLEHLDEASFSDIYSGFI 120  
Db 61 ADPQTSTILYSMLLQNFEMAAARGCWALDELHNRLCFQOSLEHLDEASFSDIYSGFI 120

QY 121 EHAAEVREYIAQLDESSAA 139  
Db 121 EHAAEVREYIAQLDESSAA 139

## RESULT 3

AAV84857  
ID AAY84857 standard; protein; 139 AA.

XX  
AC AAY84857;

XX  
DT 08-AUG-2000 (first entry)

XX  
DE A hypersensitive response elicitor protein.

XX  
KW Hypersensitive response; insect control; disease resistance;

XX  
KW Hypersensitive response elicitor; plant growth; vegetable; crop;

XX  
KW ornamental plant; dspf gene.

XX  
OS Erwinia amylovora.

XX  
PN WC200020452-A2.

XX  
PD 13-APR-2000.

XX  
PF 05-OCT-1999; 99WO-US023181.

XX  
PR 05-OCT-1998; 98US-0103050P.

XX  
PA (BDEN-) EDEN BIOSCIENCE CORP.

XX  
PI Wei Z, Niggemeyer JL;

XX  
DR WPI; 2000-303745/26.

XX  
DR N-PSDB; AAA14941.

XX  
PT Hypersensitive response elicitor polypeptides useful for imparting  
PT enhanced growth, disease resistance and insect resistance to plants,  
PT especially vegetables and ornamental flowers.

XX  
PS Disclosure; Page 22-23; 100pp; English.

XX The present sequence represents a hypersensitive response elicitor  
CC polypeptide. The polynucleotide represents the dspf gene. The  
CC specification describes hypersensitive response elicitor polypeptide  
CC fragments, which do not elicit a hypersensitive response. Instead, the  
CC proteins impart disease resistance to plants, enhance plant growth,  
CC and/or control insects. The polypeptide fragments may be used to these  
CC properties to plants. The plants which may be treated in this way include  
CC vegetables, crops and ornamental plants such as alfalfa, rice, wheat,  
CC barley, rye, cotton, sunflower, peanut, corn, potato, sweet potato, bean,  
CC pea, chicory, lettuce, endive, cabbage, brussels sprout, beet, parsnip,  
CC turnip, cauliflower, broccoli, radish, spinach, onion, garlic, eggplant,  
CC pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear,  
CC melon, citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,  
CC tomato, sorghum or sugarcane, Arabidopsis thaliana, Sainpaulia, petunia,  
CC pelargonium, poinsettia, chrysanthemum, carnation or zinnia

XX  
SQ Sequence 139 AA;

insect control which encompasses preventing direct insect damage to plant carrier tissue colonization and plant maturation; they are also used for CC by feeding injury, interfering with insect larval feeding on the plants, CC

CC preventing insects from colonising host plants and releasing phytotoxins.  
 CC Sequences of the invention also prevent subsequent disease damage to  
 CC plants resulting from insect infection. The present sequence is *Erwinia*  
 CC amylovora hypersensitive response elicitor protein, dspF

XX Sequence 139 AA;

Query Match 100.0%; Score 715; DB 5; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-78;  
 Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTSSQORVERFLOYSAGCKTPHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRRIE 60  
 Db |||||  
 QY 1 MTSSQORVERFLOYSAGCKTPHLKDGVCALYNEQDEAAVLEVPQHSLSLLHCRRIE 60  
 Db |||||  
 QY 61 ADPQTSITLYSMLLQNFENMAARGCWLDELHNVRLCFQOSLEHLDEASFSDIYSGFI 120  
 Db 61 ADPQTSITLYSMLLQNFENMAARGCWLDELHNVRLCFQOSLEHLDEASFSDIYSGFI 120  
 QY 121 EHAAEVREYIAQDESSAA 139  
 Db |||||  
 QY 121 EHAAEVREYIAQDESSAA 139

# RESULT 6

ABU20768  
 ID ABU20768 standard; protein; 792 AA.  
 AC ABU20768;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #6295.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Bacteroides fragilis.  
 XX  
 PN WO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR N-PSDB; ACA24638.

XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids, required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 48692; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for  
 CC proliferation; (7) identifying a compound that influences the activity of  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising; (12) determining the extent  
 CC product is overexpressed or underexpressed; (13) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 792 AA;

Query Match 11.5%; Score 82; DB 6; Length 792;  
 Best Local Similarity 22.5%; Pred. No. 2;  
 Matches 42; Conservative 27; Mismatches 42; Indels 76; Gaps 10;  
 QY 1 MTSSQORVER-----FLQFSAGCKTPHLKDGVCALYN----- 34  
 Db |||||  
 QY 35 ---EQDEAAVLEVPQHSLSLLHCRRIEADPQTSITLYSMLLQ-----LNFEAAM 83  
 Db 592 EIDPDEKASMLEIHNHNNELLD-----KLNDVLEISRLDSGLDFDMKE- 637  
 QY 84 RGCWLALD---ELHN-----VRLCFQOSLE-----HLDEASFSDIYSGFIEHAAEV 126  
 Db 638 ---WNMTDIVKEIYKTYKTYPLRLSLQFRLELDDTVSPVPHTRLRFVQVISFLNNANKF 694  
 QY 127 RE--YIA 131  
 Db 695 TQNGYIA 701

# RESULT 7

ABU40422  
 ID ABU40422 standard; protein; 402 AA.

XX AC

XX ABU40422;

XX AC

XX 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #25949.

XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Pseudomonas putida.

XX PN WO200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.



KW Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;  
 KW chromosome 1; chromosome 4; immune system; splice variant.  
 XX Homo sapiens.  
 XX WO200043525-A2.  
 XX 27-JUL-2000.  
 XX 21-JAN-2000; 2000WO-US001338.  
 XX 21-JAN-1999; 99US-0116670P.  
 PR 14-JUN-1999; 99US-0138682P.  
 PR 27-SEP-1999; 99US-0155798P.  
 XX (IMMV ) IMMUNEX CORP.  
 XX Cerretti DP;  
 XX WPI; 2000-482914/42.  
 DR N-PSDB; AAAS9304.  
 XX Snake venom protease (SVPH) nucleic acids, and polypeptides, used to  
 PT identify proteins having metalloproteinase-disintegrin activity, and  
 PT inhibitors of the proteins for use in therapeutics.  
 XX Claim 13; Page 15; 105pp; English.  
 XX The present sequence represents an alternatively spliced snake venom  
 CC protease-1 (SVPH-1) polypeptide, designated SVPH-1a. The SVPH  
 CC polypeptides are metalloproteinase-disintegrin protein family members.  
 CC The SVPH polynucleotides can be used as probes to identify nucleic acids  
 CC encoding proteins having metalloproteinase-disintegrin activity, to  
 CC identify human chromosome 1 or 4, to map genes on those chromosomes, to  
 CC identify genes associated with diseases, syndromes and conditions  
 CC associated with the chromosomes, and to study proteinases and their  
 CC activities on cell/cell interactions and the immune system. Sense or  
 CC antisense oligonucleotides of SVPH can be used to inhibit gene expression  
 CC of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cell/cell  
 CC and cell/matrix interactions involved in cellular processes and in the  
 CC immune system. The polypeptides may also be used to screen for inhibitors  
 CC of the polypeptide's activity, which are used in therapeutics. The  
 CC antibodies can be used in assays to detect the presence of the  
 CC polypeptides in vitro or in vivo, and to purify the polypeptides by  
 CC affinity chromatography  
 XX Sequence 766 AA;  
 SQ  
 Query Match 11.3%; Score 80.5; DB 3; Length 766;  
 Best Local Similarity 22.3%; Pred. No. 2.9;  
 Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;  
 QY 18 GCKTPHLK-----DGVCALYNEQDEAAVLEVPQHSLSLLHCRIRIADPQTSITL-- 69  
 Db 60 GQKHITHIKVKLLFSKHLFPVFTYDQGAILEDQPFVQNNCYH-GYVEGDPESLSLST 118  
 QY 70 ----YSMLLQLNFEMAMRGCLALDELHNVRLCFOQSLEHL-----DEAGSFSDIVSG 118  
 Db 119 CFGGFGQILQIN-----DPAYEIKPLAFSTTFEHLVYKMDSEKQFSTWRSG 165  
 QY 119 FIEHAAEVREYIAQLDESS 137  
 Db 166 FMQNEITCRMEFEIDNST 184  
 RESULT 10  
 AAB07740  
 ID AAB07740 standard; protein; 787 AA.  
 XX AC AAB07740;  
 XX 07-NOV-2000 (first entry)  
 XX

DE A snake venom protease (SVPH-1) polypeptide variant SVPH-1b.  
 XX Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;  
 KW chromosome 1; chromosome 4; immune system; splice variant.  
 XX Homo sapiens.  
 XX WO200043525-A2.  
 XX 27-JUL-2000.  
 XX 21-JAN-2000; 2000WO-US001338.  
 XX 21-JAN-1999; 99US-0116670P.  
 PR 14-JUN-1999; 99US-0138682P.  
 PR 27-SEP-1999; 99US-0155798P.  
 XX (IMMV ) IMMUNEX CORP.  
 XX Cerretti DP;  
 XX WPI; 2000-482914/42.  
 DR N-PSDB; AAAS9305.  
 XX Snake venom protease (SVPH) nucleic acids, and polypeptides, used to  
 PT identify proteins having metalloproteinase-disintegrin activity, and  
 PT inhibitors of the proteins for use in therapeutics.  
 XX Claim 13; Page 15-16; 105pp; English.  
 XX The present sequence represents an alternatively spliced snake venom  
 CC protease-1 (SVPH-1) polypeptide, designated SVPH-1b. The SVPH  
 CC polypeptides are metalloproteinase-disintegrin protein family members.  
 CC The SVPH polynucleotides can be used as probes to identify nucleic acids  
 CC encoding proteins having metalloproteinase-disintegrin activity, to  
 CC identify human chromosome 1 or 4, to map genes on those chromosomes, to  
 CC identify genes associated with diseases, syndromes and conditions  
 CC associated with the chromosomes, and to study proteinases and their  
 CC activities on cell/cell interactions and the immune system. Sense or  
 CC antisense oligonucleotides of SVPH can be used to inhibit gene expression  
 CC of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cell/cell  
 CC and cell/matrix interactions involved in cellular processes and in the  
 CC immune system. The polypeptides may also be used to screen for inhibitors  
 CC of the polypeptide's activity, which are used in therapeutics. The  
 CC antibodies can be used in assays to detect the presence of the  
 CC polypeptides in vitro or in vivo, and to purify the polypeptides by  
 CC affinity chromatography  
 XX Sequence 787 AA;  
 SQ  
 Query Match 11.3%; Score 80.5; DB 3; Length 787;  
 Best Local Similarity 22.3%; Pred. No. 3;  
 Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;  
 QY 18 GCKTPHLK-----DGVCALYNEQDEAAVLEVPQHSLSLLHCRIRIADPQTSITL-- 69  
 Db 60 GQKHITHIKVKLLFSKHLFPVFTYDQGAILEDQPFVQNNCYH-GYVEGDPESLSLST 118  
 QY 70 ----YSMLLQLNFEMAMRGCLALDELHNVRLCFOQSLEHL-----DEAGSFSDIVSG 118  
 Db 119 CFGGFGQILQIN-----DPAYEIKPLAFSTTFEHLVYKMDSEKQFSTWRSG 165  
 QY 119 FIEHAAEVREYIAQLDESS 137  
 Db 166 FMQNEITCRMEFEIDNST 184  
 RESULT 11  
 AAB07741  
 ID AAB07741 standard; protein; 820 AA.  
 XX AC AAB07741;  
 XX



DT	07-NOV-2000	(first entry)
DE	A snake venom protease (SVPH-1) polypeptide variant SVPH-1c.	
DE	Snake venom protease; SVPH; SVPH-1; metalloproteinase-disintegrin;	
KW	chromosome 1; chromosome 4; immune system; splice variant.	
KW	Homo sapiens.	
OS	WO200043525-A2.	
PN	27-JUL-2000.	
PD	21-JAN-2000; 200WO-US001338.	
PD	21-JAN-1999; 99US-0116670P.	
PF	14-JUN-1999; 99US-0138682P.	
XX	27-SEP-1999; 99US-0155798P.	
XX	(IMMV ) IMMUNEX CORP.	
PA	Cerretti DP;	
XX	WPI: 2000-482914/42.	
XX	N-PSDB; AAS59306.	
DR	Snake venom protease (SVPH) nucleic acids, and polypeptides, used	
DR	identify proteins having metalloproteinase-disintegrin activity, a	
DR	inhibitors of the proteins for use in therapeutics.	
PT	Claim 13; Page 16; 105pp; English.	
PT	The present sequence represents an alternatively spliced snake ven	
PT	protease-1 (SVPH-1) polypeptide, designated SVPH-1c. The SVPH	
PT	polypeptides are metalloproteinase-disintegrin protein family memb	
PT	The SVPH polynucleotides can be used as probes to identify nucleic	
PT	encoding proteins having metalloproteinase-disintegrin activity, t	
PT	identify human chromosome 1 or 4, to map genes on those chromosom	
PT	identify genes associated with diseases, syndromes and conditions	
XX	associated with the chromosomes, and to study proteinases and thei	
XX	activities on cell/cell interactions and the immune system. Sense	
XX	antisense oligonucleotides of SVPH can be used to inhibit gene exp	
XX	of SVPH 1, 3, or 4. The SVPH polypeptides can be used to study cel	
XX	and cell/matrix interactions involved in cellular processes and in	
XX	immune system. The polypeptides may also be used to screen for inh	
XX	of the polypeptide/s activity, which are used in therapeutics. The	
XX	antibodies can be used in assays to detect the presence of the	
XX	polypeptides in vitro or in vivo, and to purify the polypeptides b	
XX	affinity chromatography	
XX	Sequence 820 AA;	
XX	SQ	

AC	ADE78250;
XX	
XX	29-JAN-2004 (first entry)
XX	
DE	Endometrial specific protein identified as DEX0379_11_aa_15 (SeqID 157).
XX	
KW	neoplastic; endometrial cell; endometrial specific nucleic acid; ESNAs;
KW	cancer; metastasis; gene therapy; cytostatic; human.
XX	
OS	Homo sapiens.
XX	
PN	WO2003055982-A2.
XX	
PD	10-JUL-2003.
XX	
PF	20-DEC-2002; 2002WO-US041175.
XX	
PR	21-DEC-2001; 2001US-0343134P.
XX	
PA	(DIAD-) DIADEXUS INC.
XX	
PI	Sun Y, Liu C;
XX	
DR	WPI; 2003-569441/53.
DR	N-FSDB; ADE78347.
XX	
PT	New nucleic acid, useful for preparing a composition for diagnosing or
PT	treating endometrial cancer.
XX	
PS	Claim 1; SEQ ID NO 157; 327pp; English.
XX	
CC	This invention relates to novel nucleic acid molecules, and encoded
CC	proteins thereof, which are present in normal and neoplastic endometrial
CC	cells. Specifically, it refers to the use of these endometrial specific
CC	nucleic acids (ESNAs), as well as suitable antibodies, agonists and
CC	antagonists that are useful for the identification, diagnosis and
CC	monitoring of endometrial cancer. The present invention describes a
CC	method for monitoring the presence of an endometrial specific protein in
CC	a sample (potentially representing metastases) that comprises contacting
CC	the sample with a reagent and detecting the intensity of this
CC	interaction. Accordingly, via gene therapy, these ESNAs are useful for
CC	preparing a cytostatic composition for diagnosing or treating endometrial
CC	cancer. This polypeptide sequence is a human endometrial specific
CC	polypeptide of the invention.
XX	
QQ	Sequence 957 AA;
Query Match	11.1%; Score 79.5; DB 7; Length 957;
Best Local Similarity	32.8%; Pred. No. 5.2;
Matches	21; Conservative 12; Mismatches 16; Indels 15; Gaps 3
Qy	67 ITLYSMILQLNFEMAMRGCMWLDELHN-VRLCFQQSLEHLDEASFGSDIVSGFTEHAAE 125
Dd	354 ILLFAICLVLPET-----ILLNEIYNVLIQFQ-----EARFPDVANGFIEIIH 399  :::   :::   :::   :::   :::   :::   :::
Qy	126 VREY 129
	:
Dd	400 FKNY 403
RESULT 13	
AAG81193	ID
AAG81193	standard; protein; 789 AA.
XX	
AC	AAG81193;
XX	
DT	04-SEP-2001 (first entry)
XX	
DE	Mycobacterium tuberculosis potential drug target protein SEQ ID 244.
XX	
KW	Drug target; growth; organism viability; characterisation.
XX	
OS	Mycobacterium tuberculosis.

XX WO200135317-A1.  
 XX 17-MAY-2001.  
 XX  
 XX 13-NOV-2000; 2000WO-US031152.  
 XX  
 XX 12-NOV-1999; 99US-0165086P.  
 PR 12-NOV-1999; 99US-0165124P.  
 PR 01-FEB-2000; 2000US-0179531P.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX Eisenberg D, Rotstein SH, Marcotte EM;  
 XX  
 XX WPI; 2001-329193/34.  
 DR N-PSDB; AAH52044.  
 XX  
 XX Identifying nucleotide or polypeptide sequence for use as drug target,  
 PT involves providing algorithm that analyzes a functional relationship  
 PT between nucleotide or polypeptide sequences, and comparing the sequences.  
 XX  
 XX Disclosure; Page 178; 207pp; English.  
 XX  
 XX This invention relates to a method for identifying a nucleotide or  
 CC polypeptide sequence that may be a drug target, or essential for growth  
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092  
 CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium  
 CC tuberculosis proteins which are potential drug targets. The DNA and  
 CC protein sequences are used to illustrate the method of the invention. The  
 CC method involves providing an unknown nucleotide or polypeptide sequences,  
 CC and comparing it to a number of sequences along with at least one  
 CC algorithm capable of analysing a functional relationship between  
 CC nucleotide and polypeptide sequences. The method is useful for  
 CC characterising the function of nucleic acids and polypeptides that may be  
 CC useful as a target for a drug or essential for the growth or viability of  
 CC an organism  
 XX  
 XX Sequence 789 AA;  
 SQ  
 Query Match 11.0%; Score 79; DB 4; Length 789;  
 Best Local Similarity 24.3%; Fred. No. 4.6;  
 Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;  
 QY 23 IHLKGVKALYNEQDEAAVLEVPQSDSLHLCRIIEADPQTSITLYSMILQINTEMAA 82  
 Db 445 VSMRQYLGAPHGELTQDPKAKRLAQKMSFEVAVRILQATPVATGLVSAALL-----LT 498  
 QY 83 MRGCLALDELHNVRLCFQOSLEHLD-----BASFSDIVSGF---- 119  
 Db 499 TRGTALTLDQLHHT--LQSDLDYLRKQSPVSTSAIHLRSREGVRAAADALSGNHPVTR 555  
 QY 120 IEHAAVREYIAQLDESSAA 139  
 Db 556 VDSGRPEVWYIAPDDEHAAA 575  
 RESULT 14  
 AAW44864  
 ID AAW44864 standard; protein; 1105 AA.  
 XX  
 XX AAW44864;  
 AC  
 XX  
 XX 28-AUG-1998 (first entry)  
 DT  
 XX  
 XX Human TPC2 telomere length and telomerase regulatory protein.  
 DE  
 XX  
 XX TPC2; telomere length; telomerase; human; cancer; gene therapy;  
 KW diagnosis; vaccine.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH

FT Misc-difference 182 /note= "encoded by TTT"  
 FT FT Misc-difference 211 /note= "encoded by NGG"  
 FT FT Misc-difference 217 /note= "encoded by GCT, C being uncertain"  
 FT FT Misc-difference 231 /note= "encoded by ATC, C being uncertain"  
 FT FT Misc-difference 381 /note= "encoded by TTA, T being uncertain"  
 FT FT Misc-difference 383 /note= "encoded by AYT"  
 FT FT Misc-difference 387 /note= "encoded by GGC, the second G being uncertain"  
 FT FT Misc-difference 396 /note= "encoded by GAG, the second G being uncertain"  
 FT FT Misc-difference 403 /note= "encoded by CYT"  
 FT FT Misc-difference 404 /note= "encoded by TCA, C being uncertain"  
 FT FT Misc-difference 409 /note= "encoded by GAA, G being uncertain"  
 FT FT Misc-difference 410 /note= "encoded by CAG, G being uncertain"  
 FT FT Misc-difference 413 /note= "encoded by ACT, C being uncertain"  
 FT FT Misc-difference 416 /note= "encoded by TAC, A being uncertain"  
 FT FT Misc-difference 419 /note= "encoded by GCT"  
 FT FT Misc-difference 421 /note= "encoded by CYT"  
 FT FT Misc-difference 426 /note= "encoded by CYT"  
 FT FT Misc-difference 434 /note= "encoded by AGC, G being uncertain"  
 FT FT Misc-difference 459 /note= "encoded by CTT, C being uncertain"  
 FT FT Misc-difference 460 /note= "encoded by CCT, the second C being uncertain"  
 FT FT Misc-difference 464 /note= "encoded by CCT, the second C being uncertain"  
 FT FT Misc-difference 469 /note= "encoded by CCT, the second C being uncertain"  
 FT FT Misc-difference 476 /note= "encoded by AYT"  
 FT FT Misc-difference 477 /note= "encoded by YTT"  
 FT FT Misc-difference 487 /note= "encoded by GGY"  
 FT FT Misc-difference 488 /note= "encoded by TCT"  
 FT FT Misc-difference 491 /note= "encoded by CYT"  
 FT FT Misc-difference 510 /note= "encoded by AAG, G being uncertain"  
 FT FT Misc-difference 515 /note= "encoded by GAG, both Gs being uncertain"  
 FT FT Misc-difference 517 /note= "encoded by CGN"  
 FT FT Misc-difference 527 /note= "encoded by GYT"  
 FT FT Misc-difference 529 /note= "encoded by CGA, G being uncertain"  
 FT FT Misc-difference 530 /note= "encoded by GYG, the first G being uncertain"  
 FT FT Misc-difference 551 /note= "encoded by AGA, G being uncertain"  
 FT FT Misc-difference 562 /note= "encoded by GCT, C being uncertain"  
 FT FT Misc-difference 563 /note= "encoded by AGA, G being uncertain"  
 FT FT Misc-difference 569 /note= "encoded by AGA, G being uncertain"



Search completed: July 7, 2004, 15:17:56  
Job time : 9.80425 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:15:47 ; Search time 2.67172 Seconds  
(without alignments)  
2685.909 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....TEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	100.0	139	3	US-09-120-663-4
2	715	100.0	139	4	US-09-431-614-10
3	80.5	11.3	368	2	US-08-869-137-2
4	76.5	10.7	1105	2	US-08-710-249-2
5	76.5	10.7	1105	4	US-09-220-157A-2
6	75	10.5	243	4	US-08-858-207A-338
7	74	10.3	424	4	US-09-543-681A-7210
8	74	10.3	589	4	US-09-540-236-3398
9	73.5	10.3	984	2	US-08-673-789-9
10	73.5	10.3	984	2	US-08-449-645A-19
11	73.5	10.3	984	2	US-08-702-367A-19
12	73.5	10.3	984	5	PCT-US95-04681-19
13	73	10.2	217	4	US-09-489-039A-7789
14	73	10.2	560	4	US-09-134-001C-4343
15	71.5	10.0	572	4	US-09-543-681A-8138
16	70.5	9.9	103	4	US-09-107-532A-5377
17	70	9.8	946	4	US-09-657-931A-10
18	69.5	9.7	2183	1	US-08-348-891A-7
19	69.5	9.7	2183	2	US-08-905-817-7
20	68.5	9.6	107	4	US-09-489-039A-9798
21	68.5	9.6	239	4	US-09-485-885-12
22	68.5	9.6	591	1	US-08-179-738-5
23	68.5	9.6	591	1	US-08-179-738-10
24	68.5	9.6	591	2	US-08-628-145-5
25	68.5	9.6	591	2	US-08-628-145-10
26	68.5	9.6	595	1	US-08-171-718-16
27	68.5	9.6	595	3	US-08-478-087-16

28 68.5 9.6 596 1 US-08-179-738-2 Sequence 2, Appli  
29 68.5 9.6 596 1 US-08-179-738-3 Sequence 3, Appli  
30 68.5 9.6 596 2 US-08-628-145-2 Sequence 3, Appli  
31 68.5 9.6 596 2 US-08-628-145-3 Sequence 3, Appli  
32 68 9.5 225 4 US-09-198-452A-80 Sequence 80, Appli  
33 68 9.5 371 4 US-09-543-681A-7332 Sequence 7332, Ap  
34 68 9.5 652 1 US-08-261-663A-6 Sequence 6, Appli  
35 68 9.5 652 4 US-09-357-206A-5 Sequence 5, Appli  
36 68 9.5 652 4 US-09-813-742A-5 Sequence 5, Appli  
37 68 9.5 652 4 PCT-US95-07754A-6 Sequence 6, Appli  
38 68 9.5 1143 2 US-08-310-912A-108 Sequence 108, App  
39 68 9.5 1143 3 US-09-301-085-108 Sequence 108, App  
40 68 9.5 1143 5 PCT-US95-04589-108 Sequence 108, App  
41 68 9.5 1144 1 US-08-261-663A-2 Sequence 2, Appli  
42 68 9.5 1144 1 US-08-261-663A-4 Sequence 4, Appli  
43 68 9.5 1144 3 US-08-930-996A-9 Sequence 9, Appli  
44 68 9.5 1144 4 US-09-357-206A-3 Sequence 3, Appli  
45 68 9.5 1144 4 US-09-813-742A-3 Sequence 3, Appli

#### ALIGNMENTS

RESULT 1  
US-09-120-663-4  
; Sequence 4, Application US/09120663  
; Patent No. 6228644  
; GENERAL INFORMATION:  
; APPLICANT: Bogdanove, Adam J.  
; APPLICANT: Kim, Jihyun Francis  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Beer, Steven V.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM  
; TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
; STREET: P.O. Box 1051, Clinton Square  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/120,663  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/055,105  
; FILING DATE: 06-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Goldman, Michael L.  
; REGISTRATION NUMBER: 30,727  
; REFERENCE/DOCKET NUMBER: 19603/1661  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (716) 263-1304  
; TELEFAX: (716) 263-1600  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 139 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-09-120-663-4

Query Match 100.0%; Score 715; DB 3; Length 139;  
Best Local Similarity 100.0%; Pred. No. 4.2e-81;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTSSQORVERFLOYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSDSLHLHCRILIE 60  
Db 1 MTSSQORVERFLOYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSDSLHLHCRILIE 60  
QY 61 ADPQTSITLYSMLLQINFEMAMRGCGWLALDELHNVRCLFCQOSLEHLDEASFSDIYSGFI 120  
Db 61 ADPQTSITLYSMLLQINFEMAMRGCGWLALDELHNVRCLFCQOSLEHLDEASFSDIYSGFI 120  
QY 121 EHAAEVREYIAQLDESSAA 139  
Db 121 EHAAEVREYIAQLDESSAA 139

## RESULT 2

US-09-431-614-10  
; Sequence 10, Application US/09431614  
; Patent No. 6624139  
; GENERAL INFORMATION:  
; APPLICANT: Wei, Zhong-Min  
; APPLICANT: Schading, Richard L.  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS  
; TITLE OF INVENTION: RESISTANCE  
; FILE REFERENCE: 21829/41 (EBC-003)  
; CURRENT APPLICATION NUMBER: US/09/431,614  
; CURRENT FILING DATE: 1999-11-02  
; EARLIER APPLICATION NUMBER: 60/107,243  
; EARLIER FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patent in ver. 2.0  
; SEQ ID NO 10  
; LENGTH: 139  
; TYPE: PRT  
; ORGANISM: Erwinia amylovora  
US-09-431-614-10

Query Match 100.0%; Score 715; DB 4; Length 139;  
Best Local Similarity 100.0%; Pred. No. 4.2e-81;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MTSSQORVERFLOYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSDSLHLHCRILIE 60  
Db 1 MTSSQORVERFLOYFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSDSLHLHCRILIE 60  
QY 61 ADPQTSITLYSMLLQINFEMAMRGCGWLALDELHNVRCLFCQOSLEHLDEASFSDIYSGFI 120  
Db 61 ADPQTSITLYSMLLQINFEMAMRGCGWLALDELHNVRCLFCQOSLEHLDEASFSDIYSGFI 120  
QY 121 EHAAEVREYIAQLDESSAA 139  
Db 121 EHAAEVREYIAQLDESSAA 139

## RESULT 3

US-08-869-137-2  
; Sequence 2, Application US/08869137  
; Patent No. 5856157  
; GENERAL INFORMATION:  
; APPLICANT: Schultz, David  
; APPLICANT: Craig, Richard  
; APPLICANT: Medford, June I.  
; APPLICANT: Mumma, R.O.  
; APPLICANT: Cox-Foster, Diana L.  
; TITLE OF INVENTION: A No. 5856157el D9 14:0-ACP Fatty  
; TITLE OF INVENTION: Acid Desaturase and Gene Therefor  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Webb Law Firm  
; STREET: 700 Koppers Building, 436 Seventh Avenue  
; CITY: Pittsburgh  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 15219-1818

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/869,137  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/018,957  
FILING DATE: 04-JUN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Johnson, Barbara E  
REGISTRATION NUMBER: 31,198  
REFERENCE/DOCKET NUMBER: 2034-970766  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 412-471-8815  
TELEFAX: 412-471-4094  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 368 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-869-137-2

Query Match 11.3%; Score 80.5; DB 2; Length 368;  
Best Local Similarity 20.1%; Pred. No. 0.14; Indels 67; Gaps 8;  
Matches 37; Conservative 26; Mismatches 54; Indels 67; Gaps 8;

QY 6 QVRFELQYFSA-----GC-KTPIHLKDGVCALYNEQDEAAVLE-----VPQHSDSL 52  
Db 169 RQVEKTIQVLIALGQDIGTEKNPYHL-----FYTSFOERATFISHANTAKLAQQHGDQK 223  
QY 53 LLH-CRIEADPQTSITLYSMLLQINFEM-----FYTSFOERATFISHANTAKLAQQHGDQK 80  
Db 224 LAQICGTIAADEKRHETAYTRIVDKLFELDPDETMSCLAHMKRKITMPAHLMRDGRDPH 283  
QY 81 -----AAMRCGWALDELHNVRCLFCQOSLEH-LDEASFSDIVSGFIEHAAEVREYIA 131  
Db 284 LFQHFVSWASRTGVYTMVDYIN-----ILEHFVEKWNIEKITAGLSKGRGAQDYVC 335  
QY 132 QLDE 135  
Db 336 KLGE 339

## RESULT 4

US-08-710-249-2  
; Sequence 2, Application US/08710249  
; Patent No. 5858777  
; GENERAL INFORMATION:  
; APPLICANT: Villeponteau, Bryant  
; APPLICANT: Feng, Junli  
; APPLICANT: Adams, William H.  
; APPLICANT: Adams, Robert R.  
; TITLE OF INVENTION: Methods and Reagents for Regulating  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible



```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 338:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
; US-08-858-207A-338

Query Match 10.5%; Score 75; DB 4; Length 243;
Best Local Similarity 23.8%; Pred. No. 0.39;
Matches 29; Conservative 25; Mismatches 30; Indels 38; Gaps 7;

QY 20 KPIHLKDG--VCA-----LYNQDEEAALVFPQHSLSLLHC 56
DB 106 KPISSRDGMKLCVETSTTFDDWVROSEKDYQDMLIYLKENDFAKIGELTE-KNALAMH- 163
QY 57 RIIEADPQTSITLYSMLQLNPE-MAAMR-----GCLALDELHNRL-CFQOQSLHL 107
DB 164 ----ATTKASPAPSYLTDAASYEAMAFVRLREKGEACFTMDAGENVKVFQEKOLEHL 219
QY 108 DE 109
DB 220 SE 221

RESULT 7
US-09-543-681A-7210
; Sequence 7210, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7210
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-7210

Query Match 10.3%; Score 74; DB 4; Length 424;
Best Local Similarity 24.8%; Pred. No. 1.1;
Matches 32; Conservative 25; Mismatches 42; Indels 30; Gaps 5;

QY 25 LKDGVCALYNEQDEEAALVFPQHSLSLLH-----CRIEADPQTSITLYSMLQLNPE 79
DB 119 LINGVLRLQFQROD--VLMERFQNDSRVLPKSWLLTRIKNAYPE---LWESIIEGNNQ 172
QY 80 MAAMRCGLALDELHNRLCFQOQSLHLDEAFSD-----IVSGFIEHA 123
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 338:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6348328e
; US-08-858-207A-338

Query Match 10.3%; Score 74; DB 4; Length 589;
Best Local Similarity 23.8%; Pred. No. 1.8;
Matches 19; Conservative 17; Mismatches 26; Indels 18; Gaps 2;

QY 23 IHLKGVCAALYNEQDEEAALVFPQ-----HSDSLHCRRIEADPQTSITLYSML 73
DB 215 VEVCDVAVCSIMRPDGTGQAVILNLPATVEASTNIVADQVEYFCRHLKARPHVTISLHT-- 272
QY 74 LQINFEMAAMRGCLALDEL 93
DB 273 -----HNDRCGCAVAASEL 285

RESULT 9
US-08-673-789-9
; Sequence 9, Application US/08673789
; Patent No. 5814479
; GENERAL INFORMATION:
; APPLICANT: ZHOU, RENPING; SCHULZ, NICHOLAS,
; APPLICANT: T.; KROMER, LAWRENCE, F.; VANDE WOUDE,
; APPLICANT: GEORGE, F.
; TITLE OF INVENTION: BSK RECEPTOR LIKE
; TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR
; TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/673,789
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/177,812
; FILING DATE: 04-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: CAROL M. GRUPPI
; REGISTRATION NUMBER: 37,341

```





```

CORRESPONDENCE ADDRESS:
ADDRESSER: Amgen Patent Operations/RBW
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04681
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-287
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04681-19

Query Match 10.3%; Score 73.5; DB 5; Length 984;
Best Local Similarity 27.5%; Pred. No. 4.4; Indels 23; Gaps 6;
Matches 36; Conservative 20; Mismatches 52;

QY 7 RVERFLQYFSGCKT-----PIHLKGVCAALYNEQDEEAALVLEVPQHSLSLLHCRHIE 60
DB 94 RVHVELQFTVRDCKSPGAGPCKCKETFNLLYMSDQDVGI-----QLRPLFKQVTTVA 149
QY 61 ADPQTSI-TLYSMLLQINFMAAM-----RGCWALDE-----LHNVRLCFQOSLEHLD 108
DB 150 ADQSFTIRDLASGVKLNVERCSLGRLTRRGLVLAHPNPGACVALSVRVFYQRCPTLN 209
QY 109 E-ASFSDIVSG 118
DB 210 GLAQFFDTLFG 220

RESULT 13
US-09-489-039A-7789
; Sequence 7789, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 1432
; SEQ ID NO 7789
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7789

Query Match 10.2%; Score 73; DB 4; Length 217;
Best Local Similarity 22.1%; Pred. No. 0.58; Indels 37; Gaps 5;
Matches 33; Conservative 17; Mismatches 37;

QY 28 GVCALYNEQDEEAALVLEVPQ-----HSDLSLLHC-----RIIEADPQTSITLYS- 71
DB 2 GVCVKNYYDITLALAGVCQARLVQLAHQCHCDSDALVLSNLSIIDLPSTLAVFGG 61

CORRESPONDENCE ADDRESS:
ADDRESSER: Amgen Patent Operations/RBW
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04681
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-287
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 984 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04681-19

Query Match 10.3%; Score 73.5; DB 5; Length 984;
Best Local Similarity 27.5%; Pred. No. 4.4; Indels 23; Gaps 6;
Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6;

QY 7 RVERFLQYFSGCKT-----PIHLKGVCAALYNEQDEEAALVLEVPQHSLSLLHCRHIE 60
DB 94 RVHVELQFTVRDCKSPGAGPCKCKETFNLLYMSDQDVGI-----QLRPLFKQVTTVA 149
QY 61 ADPQTSI-TLYSMLLQINFMAAM-----RGCWALDE-----LHNVRLCFQOSLEHLD 108
DB 150 ADQSFTIRDLASGVKLNVERCSLGRLTRRGLVLAHPNPGACVALSVRVFYQRCPTLN 209
QY 109 E-ASFSDIVSG 118
DB 210 GLAQFFDTLFG 220

RESULT 13
US-09-489-039A-7789
; Sequence 7789, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 1432
; SEQ ID NO 7789
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-7789

Query Match 10.2%; Score 73; DB 4; Length 217;
Best Local Similarity 22.1%; Pred. No. 0.58; Indels 37; Gaps 5;
Matches 33; Conservative 17; Mismatches 37; Indels 37; Gaps 5;

QY 28 GVCALYNEQDEEAALVLEVPQ-----HSDLSLLHC-----RIIEADPQTSITLYS- 71
DB 2 GVCVKNYYDITLALAGVCQARLVQLAHQCHCDSDALVLSNLSIIDLPSTLAVFGG 61

```

```

QY 72 -----MLQLNFEMAMRGCGWLALDELHNVRL 98
DB 62 SEANRLRLGLETLGLVNTSSRQGLNAELRYTILSLMVLKRLAASKG---AMDTLGNRIA 118
QY 99 CFQOSLEHLDEAS-----FSDIVS 117
DB 119 GLHRLQEHFDLQSETLLSNMAGIYVDVIS 147

RESULT 14
US-09-134-001C-4343
; Sequence 4343, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4343
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4343

Query Match 10.2%; Score 73; DB 4; Length 560;
Best Local Similarity 26.7%; Pred. No. 2.3; Indels 30; Gaps 6;
Matches 31; Conservative 19; Mismatches 36;

QY 34 NEQDEEA--AVLEVPQHSLSLL-----HCRHIEADPQTSITLYS-----MLQLNFE 79
DB 276 NPEDHKAFQOSLEHLNADLIGTDPADRIGIVERDAEAGNIHYNGNQIGALLNYR 335
QY 80 MAMRGCGWLALDELHNVRLCFQ-----QSLHLEDEASFSDIVSGFIEHAAEVR 127
DB 336 IKQTEG-----LSN-RIMFQISVSGGLAKSLAQYHNVNFKEVLTGFKYIAAEIR 383

RESULT 15
US-09-543-681A-8138
; Sequence 8138, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8138
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8138

Query Match 10.0%; Score 71.5; DB 4; Length 572;
Best Local Similarity 20.9%; Pred. No. 3.6; Indels 58; Gaps 7;
Matches 41; Conservative 32; Mismatches 58;

QY 4 SQORVE-----RFLQYFSGCKTPIHLKDGVCALYNEQ-----DEEAALVLEVPQ 48
DB 125 SQEIRNLKGRFIEYFSAALNQLLEILLQDKNLFQEFQFTTSYQDGFYKYLAYYMRNEKF 184
QY 49 SDSLLHCRHIEADPQTSITLYSMLLQNFEMAMRGCGWLALD----- 91

```

```
Db 185 NNQL---NSALEASHORSI---ITWVILAIIVVTALIIICWFALRNALIRPLNSLLNIKAF 239
QY 92 -----ELH--NVRLCFQOSLEHLDEA-----SFSDIVSGFIEHA 123
Db 240 SEGDLRPDIEVHGRENEMSLASGLKHMQQELIHTVIRGVYQSTENIYNSTSEIAAGNNDLS 299
QY 124 AEVREYIAQLDESSAA 139
Db 300 SRTEEQVASLEETAAS 315
```

Search completed: July 7, 2004, 15:21:54  
Job time : 3.67172 secs

Page Blank (uspto)

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:20:22 ; Search time 6.67931 Seconds  
(without alignments)  
6477.980 Million cell updates/sec

Title: US-09-596-784-4  
Perfect score: 715  
Sequence: 1 MTSSQQRVERFLQFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	9	US-09-879-248-10
2	715	100.0	139	15	Sequence 10, Appl
3	86.5	12.1	142	12	Sequence 10, Appl
4	82	11.5	792	12	Sequence 24, Appl
5	81.5	11.4	402	12	Sequence 48692, A
6	80.5	11.3	766	12	Sequence 68346, A
7	80.5	11.3	787	12	Sequence 12, Appl
8	80.5	11.3	820	12	Sequence 13, Appl
9	80.5	11.3	820	12	Sequence 14, Appl
10	80	11.2	327	12	Sequence 1252, Ap
11	79	11.0	327	12	Sequence 46300, A
12	79	11.0	789	9	Sequence 239310,
13	78.5	11.0	418	15	Sequence 244, App
14	78.5	11.0	430	15	Sequence 14001, A
15	77	10.8	130	12	Sequence 17570, A
					Sequence 12, Appl

16	76.5	10.7	1105	11	US-09-895-606-2	Sequence 2, Appl
17	75.5	10.6	966	9	US-09-801-368-372	Sequence 372, App
18	75	10.5	317	14	US-10-166-225A-85	Sequence 85, Appl
19	74	10.3	422	12	US-10-282-122A-69060	Sequence 69060, A
20	74	10.3	443	15	US-10-369-493-15331	Sequence 15331, A
21	74	10.3	443	15	US-10-369-493-15698	Sequence 15698, A
22	74	10.3	443	15	US-10-369-493-16089	Sequence 16089, A
23	73.5	10.3	376	16	US-10-437-963-183670	Sequence 183670,
24	73.5	10.3	984	14	US-10-354-358-102	Sequence 102, App
25	73.5	10.3	984	15	US-10-116-275-147	Sequence 147, App
26	73.5	10.3	1237	15	US-10-334-143-33	Sequence 33, Appl
27	73.5	10.3	2032	10	US-09-978-244A-12	Sequence 12, Appl
28	73	10.2	343	12	US-10-425-114-49498	Sequence 49498, A
29	72.5	10.1	421	12	US-10-282-122A-69899	Sequence 69899, A
30	72	10.1	961	16	US-10-437-963-163590	Sequence 163590,
31	71.5	10.0	362	12	US-10-424-599-174145	Sequence 174145,
32	71.5	10.0	422	16	US-10-437-963-120281	Sequence 120281,
33	71.5	10.0	511	15	US-10-104-047-2339	Sequence 2339, Ap
34	71.5	10.0	619	15	US-10-369-493-9885	Sequence 9885, Ap
35	71.5	10.0	920	10	US-09-919-039-76	Sequence 76, Appl
36	71.5	10.0	967	14	US-10-197-666A-90	Sequence 90, Appl
37	71.5	10.0	1170	16	US-10-437-963-120284	Sequence 120284,
38	71.5	10.0	1591	14	US-10-197-666A-92	Sequence 92, Appl
39	71	9.9	178	16	US-10-437-963-117593	Sequence 117593,
40	71	9.9	226	16	US-10-437-963-177193	Sequence 177193,
41	70.5	9.9	583	16	US-10-437-963-186085	Sequence 186085,
42	70.5	9.9	987	9	US-09-736-969A-15	Sequence 15, Appl
43	70.5	9.9	987	9	US-09-736-960-15	Sequence 15, Appl
44	70.5	9.9	987	9	US-09-736-968A-15	Sequence 15, Appl
45	70.5	9.9	987	12	US-09-737-246-15	Sequence 15, Appl

## ALIGNMENTS

RESULT 1  
US-09-879-248-10  
; Sequence 10, Application US/09879248  
; Patent No. US20020062500A1  
; GENERAL INFORMATION:  
; APPLICANT: Fan, Hao  
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE  
; FILE REFERENCE: 21829/A1  
; CURRENT APPLICATION NUMBER: US/09/879,248  
; CURRENT FILING DATE: 2001-06-12  
; PRIOR FILING DATE: 2000-06-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; TYPE: PRT  
; LENGTH: 139  
; ORGANISM: Erwinia amylovora  
US-09-879-248-10

Query Match	100.0%;	Score 715;	DB 9;	Length 139;
Best Local Similarity	100.0%;	Pred. No. 5.2e-73;		
Matches 139;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MTSSQQRVERFLQFSAGCKTPTHLKDGVCALYNEQDEAAVLEVPCHSDSLHLCRIIE	60	
Db	1	MTSSQQRVERFLQFSAGCKTPTHLKDGVCALYNEQDEAAVLEVPCHSDSLHLCRIIE	60	
QY	61	ADPQTSITLYSMLQLNFENMAARGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI	120	
Db	61	ADPQTSITLYSMLQLNFENMAARGCWLALDELHNVRLCFQQSLEHLDEASFSDIVSGFI	120	
QY	121	EHAAREVREYIAQLDESSAA	139	
Db	121	EHAAREVREYIAQLDESSAA	139	

25 LKDGVCALYNEQDEBAVLEVPQHSLSLLHCRITIEADPQTSITLYSMILQINFEMAAMR 84  
26 LRDDGYLLWQGDKQKASLLVPSTGDALFAICTLSRVDPEHGRLLALHLNLSVPHVM 85  
85 GCWLALDELHNVRLCFQOSLEHLDEASPSDIVSGFIEHAAVREYIAQLDES 136  
86 SACLALDVEQNT-LCURYT--HDLGGNGADTLLALENAQALAEQIKQVLEN 134

RESULT 4  
US-10-282-122A-48692  
; Sequence 48692, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 48692  
; LENGTH: 792  
; TYPE: PRT  
; ORGANISM: Bacteroides fragilis  
US-10-282-122A-48692

Query Match 11.5%; Score 82; DB 12; Length 792;  
Best Local Similarity 22.5%; Pred. No. 3.8;  
Matches 42; Conservative 27; Mismatches 42; Indels 76; Gaps 10;

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEBAVLEVPQHSLSLLHCRITIEADPQTSITLYSMILQINFEMAAMR 84  
Db 535 MQNIQESVEREHLIVAKQMAEAKELQSFNNNSHEIRTEL---NAIVGFTNVLLGEGS 591  
QY 35 ---EQDEBAVLEVPQHSLSLLHCRITIEADPQTSITLYSMILQINFEMAAMR 83  
Db 592 BEIDPDEKASMLEITINHNNELL-----KLINDVLEIRSLDSGLDFDMKE- 637  
QY 84 RGCWLALD---ELHN-----VRLCFQOSLE-----HLDEASPSDIVSGFIEHAAV 126

25 LKDGVCALYNEQDEBAVLEVPQHSLSLLHCRITIEADPQTSITLYSMILQINFEMAAMR 84  
26 LRDDGYLLWQGDKQKASLLVPSTGDALFAICTLSRVDPEHGRLLALHLNLSVPHVM 85  
85 GCWLALDELHNVRLCFQOSLEHLDEASPSDIVSGFIEHAAVREYIAQLDES 136  
86 SACLALDVEQNT-LCURYT--HDLGGNGADTLLALENAQALAEQIKQVLEN 134

RESULT 4  
US-10-282-122A-48692  
; Sequence 48692, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Karl  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 48692  
; LENGTH: 792  
; TYPE: PRT  
; ORGANISM: Bacteroides fragilis  
US-10-282-122A-48692

Query Match 11.5%; Score 82; DB 12; Length 792;  
Best Local Similarity 22.5%; Pred. No. 3.8;  
Matches 42; Conservative 27; Mismatches 42; Indels 76; Gaps 10;

QY 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEBAVLEVPQHSLSLLHCRITIEADPQTSITLYSMILQINFEMAAMR 84  
Db 535 MQNIQESVEREHLIVAKQMAEAKELQSFNNNSHEIRTEL---NAIVGFTNVLLGEGS 591  
QY 35 ---EQDEBAVLEVPQHSLSLLHCRITIEADPQTSITLYSMILQINFEMAAMR 83  
Db 592 BEIDPDEKASMLEITINHNNELL-----KLINDVLEIRSLDSGLDFDMKE- 637  
QY 84 RGCWLALD---ELHN-----VRLCFQOSLE-----HLDEASPSDIVSGFIEHAAV 126

Query Match 12.1%; Score 86.5; DB 12; Length 142;  
Best Local Similarity 25.9%; Pred. No. 0.12;  
Matches 29; Conservative 23; Mismatches 57; Indels 3; Gaps 2;

Db 638 ---WNMTDIVKEIKTYQPLIRLSQRLLEDDTVSPVHTDLRFPVQVSNFLNNKPF 694  
QY 127 RE--YIA 131  
Db 695 TQGYIA 701

## RESULT 5

US-10-282-122A-68346  
; Sequence 68346, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 68346  
; LENGTH: 402  
; TYPE: PRT  
; ORGANISM: Pseudomonas putida

US-10-282-122A-68346

Query Match 11.4%; Score 81.5; DB 12; Length 402;  
Best Local Similarity 20.5%; Pred. No. 1.8;  
Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;  
QY 12 LQVFSACKTPI--HLKDGVCALYNEQDEEAALVLEVPQHS- 50  
Db 192 IERISRDARVPIVHKL-DGICHY-----VSQHADLDKAWNAFNAKTYRYGIC 239  
QY 51 ----SLLLHCRILIEADPQTSITLYSMLLQLNFENAMRG- 89  
Db 240 GAMETLLVDQVABR-----FLPEMARRFVEKGVLELGCRTQAIISAKPATEADWHT 293  
QY 90 -LDELHNVRIC--PQSLHLEHDEASFSDIVSGFTHEAAAEVREYIAQLDESS 137  
Db 294 YLDAILSIRVVDGLNQLAIEHNHYGSHHTDSIISEHGEARQFMAEVDSDAS 344

## RESULT 6

US-10-664-456-12  
; Sequence 12, Application US/10664456  
; Publication No. US20040038364A1  
; GENERAL INFORMATION:

; APPLICANT: Immunex Corporation  
; TITLE OF INVENTION: NEW METALLOPROTEINASE-DISINTEGRIN FAMILY MEMBERS: SVPH  
; FILE REFERENCE: 03260.0093-00304  
; CURRENT APPLICATION NUMBER: US/10/664,456  
; CURRENT FILING DATE: 2003-09-19  
; PRIOR APPLICATION NUMBER: Prior Application Number: US/09/890,323  
; PRIOR FILING DATE: Prior Filing Date: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/116,670  
; PRIOR FILING DATE: 1999-01-21  
; PRIOR APPLICATION NUMBER: 60/138,682  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: 60/155,798  
; PRIOR FILING DATE: 1999-09-27  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 12  
; LENGTH: 766  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-664-456-12

## Query Match

Best Local Similarity 11.3%; Score 80.5; DB 12; Length 766;  
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;  
QY 18 GCKTPIHLK-----DGVCALYNEQDEEAALVLEVPQHS- 69  
Db 60 GQKHLIHKVKKLLFSKHLPVFTYDQGAILEDQFVQNNCYH-GYVEGDPESLVSJST 118  
QY 70 ----YSMLLQLNFENAMRGCMWLALDELHNVR-LCFOQSLEHL-----DEASFSDIVSG 118  
Db 119 CFGGFQGIQLN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165  
QY 119 FIEHAAEVREYIAQLDESS 137  
Db 166 FMQNEITCRMEFEIDNST 184

## RESULT 7

US-10-664-456-13  
; Sequence 13, Application US/10664456  
; Publication No. US20040038364A1  
; GENERAL INFORMATION:  
; APPLICANT: Immunex Corporation  
; TITLE OF INVENTION: NEW METALLOPROTEINASE-DISINTEGRIN FAMILY MEMBERS: SVPH  
; FILE REFERENCE: 03260.0093-00304  
; CURRENT APPLICATION NUMBER: US/10/664,456  
; CURRENT FILING DATE: 2003-09-19

; PRIOR APPLICATION NUMBER: Prior Application Number: US/09/890,323  
; PRIOR FILING DATE: Prior Filing Date: 2001-07-25  
; PRIOR APPLICATION NUMBER: 60/116,670  
; PRIOR FILING DATE: 1999-01-21  
; PRIOR APPLICATION NUMBER: 60/138,682  
; PRIOR FILING DATE: 1999-06-14  
; PRIOR APPLICATION NUMBER: 60/155,798  
; PRIOR FILING DATE: 1999-09-27  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 787  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-664-456-13

## Query Match

11.3%; Score 80.5; DB 12; Length 787;

```

Best Local Similarity 22.3%; Pred. No. 5.6;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRRIIEADPQTSITL-- 69
Db 60 GQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLSLST 118
QY 70 -----YSMLLQINFEMAAMRGCLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
Db 119 CFGGFGILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165

QY 119 FIEHAAEVREYIAQLDESS 137
Db 166 FMQNEITCRMEFEIDNST 184

```

```

RESULT 8
US-10-664-456-14
; Sequence 14, Application US/10664456
; Publication No. US20040038364A1
; GENERAL INFORMATION:
; APPLICANT: Immunex Corporation
; TITLE OF INVENTION: NEW METALLOPROTEINASE-DISINTEGRIN FAMILY MEMBERS: SVPH
; TITLE OF INVENTION: DNAS AND POLYPEPTIDES
; FILE REFERENCE: 03260.0093-00304
; CURRENT APPLICATION NUMBER: US/10/664,456
; CURRENT FILING DATE: 2003-09-19
; PRIOR APPLICATION NUMBER: prior APPLICATION NUMBER: US/09/890,323
; PRIOR FILING DATE: prior FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/116,670
; PRIOR FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: 60/138,682
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: 60/155,798
; PRIOR FILING DATE: 1999-09-27
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 820
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-664-456-14

```

```

Query Match 11.3%; Score 80.5; DB 12; Length 820;
Best Local Similarity 22.3%; Pred. No. 5.9;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRRIIEADPQTSITL-- 69
Db 60 GQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLSLST 118
QY 70 -----YSMLLQINFEMAAMRGCLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
Db 119 CFGGFGILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165

QY 119 FIEHAAEVREYIAQLDESS 137
Db 166 FMQNEITCRMEFEIDNST 184

```

```

RESULT 9
US-10-408-765A-1252
; Sequence 1252, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

```

```

; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1252
; LENGTH: 820
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1252

```

```

Query Match 11.3%; Score 80.5; DB 16; Length 820;
Best Local Similarity 22.3%; Pred. No. 5.9;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPIHLK-----DGVCALYNEQDEEAAVLEVPQHSDSLLLHCRRIIEADPQTSITL-- 69
Db 60 GQKHIIHKVKLLFSKHLPVFTYTDQGAILEDQPFVQNNCYH-GYVEGDPESLVSLSLST 118
QY 70 -----YSMLLQINFEMAAMRGCLALDELHNVR-LCFQQSLEHL-----DEASFSDIVSG 118
Db 119 CFGGFGILQIN-----DFAYEIKPLAFSTTFEHLVYKMDSEKQFSTMRSG 165

QY 119 FIEHAAEVREYIAQLDESS 137
Db 166 FMQNEITCRMEFEIDNST 184

```

```

RESULT 10
US-10-425-114-46300
; Sequence 46300, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46300
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 700764472_FLI.pep
US-10-425-114-46300

```

```

Query Match 11.2%; Score 80; DB 12; Length 327;
Best Local Similarity 30.8%; Pred. No. 2;
Matches 44; Conservative 21; Mismatches 48; Indels 30; Gaps 10;

QY 24 HLKDGVCALYNEQDEE-----AAV--LEVPQHSDSLLLHCRRIIEADPQT---SIT 68
Db 16 YLENGSVEIYSRHAERTGKFPDVVAASRLKKPTVS-SLILDCIIVAYDQTQTIHSFQ 74
QY 69 LYSMLLQINFEMAAMR--GCWLALDELH-NVRLCFQQS-----BHLDEASFSDIVSGFIE 121
Db 75 ALSTRARKNVEMEDIKVDVCIFAFDLLYNGQALIQENLRVRREHL-VASFEE-EPGFLO 132

QY 122 HAA-----EVREYIAQLDESSAA 139
Db 133 FATTITSDNVEIEOKFLDQAVCA 155

```

```

RESULT 11
US-10-424-599-239310

```



```

; Sequence 239310, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 239310
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58122C.1.pep
US-10-599-239310

```

```

Query Match 11.0%; Score 79; DB 12; Length 327;
Best Local Similarity 30.8%; Pred. No. 2.6;
Matches 44; Conservative 21; Mismatches 48; Indels 30; Gaps 10;

Qy 24 HLKGVCAALYNEQDEE-----AAV--LEVPOHSDSLHLCRIIADPQT---SIT 68
Db 16 YLENGSVBIYRHAERNTGKFPVVAAVSRLLKPTVS-SLILDCIIVAYDRQTQTIHSFQ 74
Qy 69 LYSMLQLNFMAWR--GWLALDELH-NVRLCFQOSL-----EHLDEASFSDIVSGFIE 121
Db 75 ALSTRARKNVEDIKVDVCFADFLLYNGALLQENLRVRREHL-YASPEE-EPGFLQ 132
Qy 122 HAA-----EVREYIAQLDESSAA 139
Db 133 PATITSDVVEIQNLFDAQVA 155

```

```

RESULT 12
US-09-712-363-244
; Sequence 244, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 789

```

```

; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-244

Query Match 11.0%; Score 79; DB 9; Length 789;
Best Local Similarity 24.3%; Pred. No. 8.3;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

Qy 23 IHLKGVCAALYNEQDEEAAVLEVPQHSDSLHLCRIIADPQTSTITLYSMILLQNFEMAA 82
Db 445 VSMQYLGAPGELTQDPAAKRLALQKMSFEVAMRILQATPTATGLVSALL-----LT 498
Qy 83 MRGWLALDELHNVRLCFQOSLEHL-----EASFSDIVSGF----- 119
Db 499 TRGTALTLDQLHHT---LQDSLTYLERKQSPVSTSLALRSREGVRAAADALSNGHPVTR 555
Qy 120 IEHAAEVREYIAQLDESSAA 139
Db 556 VDSGREPVWYIAPDDEHAAA 575

```

```

RESULT 13
US-10-369-493-14001
; Sequence 14001, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14001
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-14001

```

```

Query Match 11.0%; Score 78.5; DB 15; Length 418;
Best Local Similarity 20.5%; Pred. No. 4.1;
Matches 35; Conservative 29; Mismatches 44; Indels 63; Gaps 8;

Qy 12 LQYFSAGCKTPI--HLKGVCAALYNEQDEEAAVLEVPQHSD----- 50
Db 213 IERISRDARVPVVKHL-DGICHVY-----VSEHADLPKAQRTAFNAKTYRYGIC 260
Qy 51 ----SLHLCRIIADPQTSTITLYSMILLQNFEMAAVRGC-----WLA- 89
Db 261 GAMETLLVDQFVAKD-----FLPSMAKQFEKGVELRGCEGTRAIIEAATAEDWSTE 314
Qy 90 -LDELHNVRLC--FQOSLEHLDEASFSDIVSGFIEHAAEVREYIAQLDESS 137
Db 315 YLAPILSIRVVDGLDQAIIEHINHFSGSHHTSDIVSENLAIDTRQFVAQVDSAS 365

```

```

RESULT 14
US-10-369-493-17570
; Sequence 17570, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```

Search completed: July 7, 2004, 15:30:45  
Job time : 7.67931 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:14:07 ; Search time 2.39049 Seconds  
(without alignments)  
5593.254 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715	100.0	139	2 T18449	pathogenicity fact
2	273.5	38.3	129	2 T30333	avirulence protein
3	79	11.0	789	2 A70868	probable plSB2 - M
4	78.5	11.0	830	2 C82710	ribonucleoside-dip
5	77.5	10.8	775	2 T45238	probable transfera
6	75.5	10.6	402	2 T19390	hypothetical prote
7	75.5	10.6	966	2 S25365	CYC8 protein - Yea
8	75	10.5	317	2 D95044	diphosphomevalonat
9	73.5	10.3	984	1 A34076	protein-tyrosine k
10	73	10.2	1708	2 AB1866	WD-40 repeat prote
11	72.5	10.1	131	2 AG2034	hypothetical prote
12	71.5	10.0	1036	2 T31673	N-acetylglucosamin
13	71.5	10.0	2484	2 T26216	hypothetical prote
14	71.5	10.0	2607	2 T26215	hypothetical prote
15	71	9.9	637	2 B95878	probable adenylate
16	70.5	9.9	1048	2 T31653	hypothetical prote
17	69.5	9.7	291	2 T23051	hypothetical prote
18	69.5	9.7	393	2 S76366	hypothetical prote
19	69.5	9.7	2183	1 ZLN2MV	genome polyprotein
20	69.5	9.7	2183	1 G48556	genome polyprotein
21	69	9.7	225	2 F72124	pes iia protein +
22	69	9.7	344	2 C97914	diphosphomevalonat
23	69	9.7	374	2 A42264	membrane-associate
24	69	9.7	451	1 F64155	hypothetical prote
25	69	9.7	610	2 T02298	probable outer den
26	69	9.7	638	2 T09400	outer dense fiber
27	69	9.7	727	2 C84748	hypothetical prote
28	68.5	9.6	130	2 AG2226	hypothetical prote
29	68.5	9.6	165	2 C72625	hypothetical prote

30 68.5 9.6 167 2 H81717 conserved hypothet  
31 68.5 9.6 173 2 S75578 hypothetical prote  
32 68.5 9.6 352 2 T44968 gas-vesicle operon  
33 68.5 9.6 591 2 I54368 merlin protein - m  
34 68.5 9.6 595 2 S33809 neurofibromin 2 -  
35 68.5 9.6 596 2 T68664 merlin - mouse  
36 68.5 9.6 621 2 T06717 hypothetical prote  
37 68.5 9.6 794 2 T46073 hypothetical prote  
38 68.5 9.6 883 2 A96805 hypothetical prote  
39 68 9.5 178 2 T39722 ORF12 - Agrobacter  
40 68 9.5 225 2 F86498 Pts IIA protein wi  
41 68 9.5 478 2 G81679 3-dehydroquinat d  
42 68 9.5 638 2 T03791 outer dense fiber  
43 68 9.5 923 2 T24712 hypothetical prote  
44 68 9.5 1119 2 B70126 surface-located me  
45 68 9.5 1144 2 A54810 TMV resistance pro

#### ALIGNMENTS

##### RESULT 1

T18449

Pathogenicity factor DspB - Erwinia amylovora

C;Species: Erwinia amylovora

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000

C;Accession: T18449

R;Gaudriault, S.

submitted to the EMBL Data Library, May 1998

A;Reference number: Z18936

A;Accession: T18449

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-139 <GAU>

A;Cross-references: EMBL:Y13831; PIDN:CAA74157.1

A;Experimental source: strain CfBPl430; specific host Pommoideae

C;Genetics:

A;Note: dspB

C;Function:

A;Description: involved in pathogenicity

Query Match 100.0%; Score 715; DB 2; Length 139;  
Best Local Similarity 100.0%; Pred. No. 4.8e-64;  
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAALVLEVPQHSLLHCRITIE 60  
|||||  
Db 1 MTSSQQRVERFLQYFSAGCKTPIHLKDGVCALYNEQDEEAALVLEVPQHSLLHCRITIE 60

Qy 61 ADPQTSITLYSMILQLINFEMAARGCWALDELHNVLRCFQQSLEHLDEASFSDIVSGFI 120  
|||||  
Db 61 ADPQTSITLYSMILQLINFEMAARGCWALDELHNVLRCFQQSLEHLDEASFSDIVSGFI 120

Qy 121 EHAAEVREYIAQLDESSAA 139  
|||||  
Db 121 EHAAEVREYIAQLDESSAA 139

##### RESULT 2

T30333

avirulence protein - Pseudomonas syringae

C;Species: Pseudomonas syringae

C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999

C;Accession: T30333

R;Bogdanov, A.J.; Kim, J.P.; Wei, Z.; Kolchinsky, P.; Charkowski, A.O.; Conlin, A.K.; Co

Proc. Natl. Acad. Sci. U.S.A. 95, 1325-1330, 1998

A;Title: Homology and functional similarity of an hrp-linked pathogenicity locus, dspeF,

A;Reference number: Z20825; MUID:98115919; PMID:9448330

A;Accession: T30333

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-129 <BOG>

A;Cross-references: EMBL:U97505; NID:g2978502; PID:g2978504; PIDN:AAC06135.1

C;Genetics:  
A:Gene: avr

	Query Match	38.3%	Score 273.5;	DB 2;	Length 129;
	Best Local Similarity	43.1%	Pred. No. 4.le-20;		
	Matches	56;	Conservative 27;	Mismatches 44;	Indels 3; Gaps 3;
Qy	1 MTSSQQRVERELQVFSACKTPHILKGVGCALYNEODEEAAVLVEPOHSDSLLLHCRIIE	60			
Dd	1 MKTSQDPFARINSIGALGTSLTLONGVCALYGQNNEAIIELPEHSEWVFCHRTGR	60			
Qy	61 ADPOTSITLYSMLQLNFEMAAMRGCMWLDELHNVRLCFOOSLEHLDEASFDIVSGFI	120			
Dd	61 C-PRAPDLL-RLLISINFDVARLHCCWFVDQ-GDVRLCQAORELASLDPEAFCDVTGCFI	117			
Qy	121 EHAAEVREYI	130			
Dd	118 SCAREARFL	127			

RESULT 3  
A70868  
probable plisB2 - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 12-Jun-2003  
C:Accession: A70868  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: A70868  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-789 <COL>  
A:Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAAL6059.1; PID:el23760  
A:Experimental source: strain H37RV  
C:Genetics:  
A:Gene: plisB2  
C:Superfamily: glycerol-3-phosphate O-acyltransferase

```

Query Match      11.0%; Score 79; DB 2; Length 789;
Best Local Similarity 24.3%; Pred. No. 8.7;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

QY 23 IHLKDGVCALNEODEEAAVLEVPQHSDSLHLHCRIIEADPQTSITYLSMLQLNFEMAA 82
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 445 VSMRQYLGAHPGHELTQDPAAKRLALQKMSFEVAVRIILQATPVTATGLVSALL -----LT 498
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 83 MRGCWLALDELHNVRLCFQOGLHL -----EASFSDIVSGF --- 119
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 499 TRGTALTUDQLHHT ---LQSDLDYLERKQSPVSTSAURLRSREGVRAAADALNSGHPVTR 555
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

QY 120 IEHAAEVREYIAQLDESSAA 139
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 556 VDSGREPVWYIAPDDEHAAA 575
      :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::

```

RESULT 4  
C82710  
ribonucleoside-diphosphate reductase alpha chain XF1196 [imported] - Xylella fastidiosa  
C/Species: Xylella fastidiosa  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000  
C/Accession: C82710  
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A/Title: The genome sequence of the plant pathogen Xylella fastidiosa.  
A/Reference number: 82515; MUID:20365717; PMID:10910347  
A/Note: for a complete list of authors see reference number A59328 below  
A/Accession: C82710  
A/Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-830 <STM>  
A:Cross-references: GB:AE003954; GB:AE003849; MID:g9106145; PIDN:AAF84006.1; GSPDB:GN001001  
A:Experimental source: strain 9a5c  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Alvarez, R.; Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, F.; Cavaletto, E.; Docena, C.; El-Doory, H.; Facincani, A.P.; Ferreira, A.J.S.  
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, M.C.; Fromhm, J.D.; Junqueira, M.L.; Kemper, E.L.; Kitaajina, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigret, E.; Machado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E.; Chado, M.A.; Oliveira, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.; Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak, A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira, M.; Tshukano, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Zamboni, A.  
A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
C:Gene: XF1196

C; Gene: *ari90*  
C; Superfamily: herpesvirus ribonucleoside-diphosphate reductase large chain

	Query Match	11.0%	Score	78.5;	DB	2;	Length	830;		
	Best Local Similarity	25.6%;	Pred. No.	10;						
	Matches	33;	Conservative	21;	Mismatches	46;	Indels	29;	Gaps	6;
QY	17	AGCKTPHLKDGVCALNEQDEEAAVLEVPQHSDSLLLHCRIL-EADPQTSTLYSMLQ	75							
			:	:	:	:	:	:	:	:
			:	:	:	:	:	:	:	:
Db	22	AGSVPPVALS-----VFHPDEFD---EVPIPQTAMMTRAVEANVTWITKEAGNR	73							
			:	:	:	:	:	:	:	:
QY	76	INFEMAAMRGCLALDLBHLNRLFCQSLBHLDEASFSDIVSGFIEHAEE-	126							
			:	:	:	:	:	:	:	:
			:	:	:	:	:	:	:	:
Db	74	MPFFCARLE---CAIDTIH-----QEFPQLDVAEYKRAVFGFVERKDSVNADDLVLL	123							
			:	:	:	:	:	:	:	:
QY	127	-REYIAQLD	134							
			:	:	:	:	:	:	:	:
Db	124	IREFEARVD	132							
			:	:	:	:	:	:	:	:

RESULT 5  
T45238  
probable transferase [imported] - Mycobacterium leprae  
C:Species: Mycobacterium leprae  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 12-Jun-2003  
C:Accession: T45238  
R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: Z22949  
A:Accession: T45238  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-775 <JAM>  
A:Cross-references: EMBL:AL049913; PIDN:CAR43153.1  
A:Experimental source: cosmid B1610  
C:Genetics:  
A:Note: plsB  
C:Superfamily: glycerol-3-phosphate O-acetyltransferase

```

Query Match      10.8%; Score 77.5; DB 2; Length 775;
Best Local Similarity 25.8%; Pred. No. 12;
Matches 34; Conservative 16; Mismatches 33; Indels 49; Gaps 5;

QY      11  FLOYFSAGCKTPIHLKDVCALYN-----EQ 36
      :  :  :  :  :  :  :  :  :  :
DB      397  YAAVARGGEKTP----EGVAMLYSFIKAQERNVGKIYVRPPEAVSMROYLGAPHGALVQ 452

QY      37  DEEAALVLEVPQHSLLILHCKRIEADPQTSITLYSMLLQLNFMMAAMRGCGWLALDELHNV 96
      :  :  :  :  :  :  :  :  :  :
DB      453  DODAKRLALQKMSFEVA--WRLICATPTATALYSALL-----LTTRGVALTDLQLHHT 504

QY      97  RLCTFQQSLEHLUD 108
      :  :  :  :  :  :  :  :  :
DB      505  ---LQESLDYLE 513

```



Db 238 -----ATTKTAPAFSYLTASAEAMAFVQLRKEAGECYFTMDAGPNVKVFCQEKDLEHL 293  
QY 108 DE 109  
Db 294 SE 295

RESULT 9  
A34076  
N;Alternate names: protein-tyrosine kinase (EC 2.7.1.112) receptor type eph 1 precursor - human  
C;Species: Homo sapiens (man)  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 04-Feb-2000  
C;Accession: A34076; S44280  
R;Hirai, H.; Maru, Y.; Hagiwara, K.; Nishida, J.; Takaku, F.  
Science 238, 1717-1720, 1987  
A;Title: A novel putative tyrosine kinase receptor encoded by the eph gene.  
A;Reference number: A34076; MUID:88070650; PMID:2825356  
A;Accession: A34076  
A;Molecule type: mRNA  
A;Residues: 1-984 <HIR>  
A;Cross-references: GB:M18391; NID:G3397116; PIDN:AAA36747.1; PID:G339717  
A;Note: the sequence in GenBank entry HUMTKR, release 111.0, has the codons GCG for 398-  
R;Tuzi, N.L.  
submitted to the EMBL Data Library, November 1993  
A;Description: An EGFR/eph chimeric receptor possesses ligand stimulated tyrosine kinase  
A;Reference number: S44280  
A;Accession: S44280  
A;Molecule type: mRNA  
A;Residues: 286-397, A', 399-580, 'QRDRATVDREKWLKPYVDLQAYEDPAQALDF', 583,625-984 <TUZ>  
A;Cross-references: EMBL:Z27409; NID:G482916; PIDN:CAA81796.1; PID:G482917  
C;Genetics:  
A;Gene: GDB:EPHT1; EPH; EPHT  
A;Cross-references: GDB:119875; OMIM:179610  
A;Map position: 7q32-7q36  
C;Superfamily: protein-tyrosine kinase, receptor type eph; fibronectin type III repeat H  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;24-984/Product: protein-tyrosine kinase receptor type eph 1 #status predicted <MAT>  
F;548-568/Domain: transmembrane #status predicted <TM>  
F;630-895/Domain: protein kinase homology <KIN>  
F;638-646/Region: protein kinase ATP-binding motif  
F;918-984/Domain: SAM homology <SAM>  
F;59,338,414,478/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 10.3%; Score 73.5; DB 1; Length 984;  
Best Local Similarity 27.5%; Pred. No. 40;  
Matches 36; Conservative 20; Mismatches 52; Indels 23; Gaps 6;

QY 7 RVERFLQYFAGCKT-----PIHLKDGVCALYNEQDEAAVLEVPQHSDSLLLHCRHIE 60  
Db 94 RVHVELQFTVRDCKSPFGAGPLGCKETFNLLYMSDQDVGI---QLRPLFKQVTTVA 149

QY 61 ADPQTSI-TLYSMLLQINFEMAM-----RCGWLADDE-----LHNVRLCFQOSLEHLD 108  
Db 150 ADQSFTRDLASGVKLVNRCVSLGRLTRGLYLAHPGACVALSVRVFYQRCPTLN 209

QY 109 E-ASFSFSDIVSG 118  
Db 210 GLAQFPDTLPG 220

RESULT 10  
AE1866  
WD-40 repeat protein [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AE1866  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AE1866  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1708 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA072436.1; PID:gl7129823; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all0478

Query Match 10.2%; Score 73; DB 2; Length 1708;  
Best Local Similarity 30.2%; Pred. No. 87;  
Matches 26; Conservative 11; Mismatches 35; Indels 14; Gaps 2;

QY 59 IEADPQTSITLYSMLLQINFEMAMRGCVLALDELHNV-----RLCFQOSLEHLD 109  
Db 276 VRPTFLSSPIRLQGVLLQYLGL-----CYCLAEQNQLDNRHWHWTAKFYQECLEILQV 330

QY 110 ASFSFSDIVSGFIEHAAEVREYIAQLDE 135  
Db 331 AGRPDIASEFIGQLAEVLEHLQAWDE 356

## RESULT 11

AG2034  
hypothetical protein all1829 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp. PCC 7120  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002  
C;Accession: AG2034  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AG2034  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-131 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA073528.1; PID:gl7130919; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: all1829

Query Match 10.1%; Score 72.5; DB 2; Length 131;  
Best Local Similarity 23.6%; Pred. No. 4.4;  
Matches 25; Conservative 20; Mismatches 46; Indels 15; Gaps 3;

QY 24 HLKDG---VCALYNEQDEAAVLEVPQHSDSLLLHCRHIEADPQTSITLYSMLLQINFEM 80  
Db 31 HLEQSDYKVCYGVWDEQDKYKEKITLPRPLETELVSSSI-----GVTHTERFLQKFSL 83

QY 81 AAMRGCVLALDELHNV-----LCFQOSLEHLDASFSFSDIVSGFIE 121  
Db 84 RARFTYIAIQANNHTQKIGELVLVYDENLEFIDENWLLDVDSMPLE 129

## RESULT 12

T31673  
N-acetylglucosaminyltransferases (EC 2.4.1.-), chain p110 - rat  
N;Alternate names: O-GlcNAc transferase  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 07-Dec-1999  
C;Accession: T31673  
R;Kreppel, L.K.; Blomberg, M.A.; Hart, G.W.  
J. Biol. Chem. 272, 9308-9315, 1997  
A;Title: Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characteri  
A;Reference number: Z21056; MUID:97238869; PMID:9083067  
A;Accession: T31673  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1036 <KRE>

A;Cross-references: EMBL:U76557; NID:g1931578; PID:g1931579; PIDN:AAC53121.1  
A;Experimental source: strain Sprague Dawley; liver  
C;Genetics:  
A;Gene: OGT  
C;Keywords: glycosyltransferase; hexosyltransferase; tandem repeat

Query Match 10.0%; Score 71.5; DB 2; Length 1036;  
Best Local Similarity 24.4%; Pred. No. 67;  
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

Qy 24 HLKDGVCALYNEQDEEAALVLEVPQHSDSLLLHCRILIEADPQT 65  
Db 281 HFDPAYCNLANALKEGSAEADCYNTALRLCPHADSLLNLANIKREQNIIEAVRL 340  
Qy 71 SMLQLNFMAAMRGWGLALDELHNVRCLFCQSLHLEHDEA-----SFSDIVSGFIEHAAE 125  
Db 341 RKALEVPEFAAHS---NLASVLQOQKLOEALMHYKEAIRISPTADAYSNMGNTLKE 397  
Qy 126 VRE 128  
Db 398 MQD 400

RESULT 13  
T26216  
hypothetical protein W06A7.3c - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T26216  
R;Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A;Reference number: Z20173  
A;Accession: T26216  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2484 <MIL>  
A;Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GN000023; CESP:W06A7.3c  
A;Experimental source: clone W06A7  
C;Genetics:  
A;Gene: CESP:W06A7.3c  
A;Map position: 5  
A;Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

Query Match 10.0%; Score 71.5; DB 2; Length 2484;  
Best Local Similarity 24.2%; Pred. No. 1.9e+02;  
Matches 31; Conservative 19; Mismatches 43; Indels 35; Gaps 6;

Qy 18 GCKTPHLKDGVCALYNEQDEEA-----AVLEVPQHSDSLLLHCRILIEADPQT 65  
Db 1491 GCITDVSAD-----VNEQDEESTLKILKWPSPSLELDFNDPKVHVPIPLMEPAT 1545  
Qy 66 SITLYSMLQLNFMAAMRGWGLALDELHNVRCLFCQSLHLEHDEASFSDIVSGFIEHAAE 125  
Db 1546 -----MYLE-----WIIADAVKEV-----SEMEVVTESEISEMAPQVSESTCP 1587  
Qy 126 VREYIAQL 133  
Db 1588 IPEPLADL 1595

RESULT 14  
T26215  
hypothetical protein W06A7.3a - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C;Accession: T26215  
R;Ainscough, R.  
submitted to the EMBL Data Library, August 1996  
A;Reference number: Z20173  
A;Accession: T26215  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2607 <MIL>

A;Cross-references: EMBL:Z78066; PIDN:CAB01522.2; GSPDB:GN000023; CESP:W06A7.3a  
A;Experimental source: clone W06A7  
C;Genetics:  
A;Gene: CESP:W06A7.3a  
A;Map position: 5  
A;Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2

Query Match 10.0%; Score 71.5; DB 2; Length 2607;  
Best Local Similarity 24.2%; Pred. No. 2e+02;  
Matches 31; Conservative 19; Mismatches 43; Indels 35; Gaps 6;

Qy 18 GCKTPHLKDGVCALYNEQDEEA-----AVLEVPQHSDSLLLHCRILIEADPQT 65  
Db 1491 GCITDVSAD-----VNEQDEESTLKILKWPSPSLELDFNDPKVHVPIPLMEPAT 1545  
Qy 66 SITLYSMLQLNFMAAMRGWGLALDELHNVRCLFCQSLHLEHDEASFSDIVSGFIEHAAE 125  
Db 1546 -----MYLE-----WIIADAVKEV-----SEMEVVTESEISEMAPQVSESTCP 1587  
Qy 126 VREYIAQL 133  
Db 1588 IPEPLADL 1595

RESULT 15  
B95878  
probable adenylate cyclase (EC 4.6.1.1) [imported] - Sinorhizobium meliloti (strain 1021)  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 14-Sep-2001  
C;Accession: B95878  
R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernanc  
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
A;Title: The complete sequence of the 1,683-Kb pSymB megaplasmid from the N2-fixing endo  
A;Reference number: A95842; MUID:21396508; PMID:11481431  
A;Accession: B95878  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-637 <KUR>  
A;Cross-references: GB:AL591985; PIDN:CAC48690.1; PID:g15140162; GSPDB:GN00167  
A;Experimental source: strain 1021, megaplasmid pSymB  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: cyaf7; SM520300  
A;Genome: plasmid  
C;Keywords: phosphorus-oxygen lyase

Query Match 9.9%; Score 71; DB 2; Length 637;  
Best Local Similarity 24.0%; Pred. No. 42;  
Matches 29; Conservative 14; Mismatches 36; Indels 42; Gaps 5;

Qy 31 ALY-----NEQDE--EAAVLEVPQHSDSLLLH-----CRIIEADPQTS 66  
Db 534 ALYAAGRLDEADVLRECLIRAPQADCLLIRAVLSORGVDVEGAQRTWARLIEVDPE-- 591  
Qy 67 ITLYSMLQLNFMAAMRGWGLALDELHNVRCLFCQSLHLEHDEASFSDIVSGFIEHAAE 126  
Db 592 -----FSLASERSMRFGDS-----ALMEQFLSLRAEANAPDVTSGFLHPTQS 635  
Qy 127 R 127  
Db 636 R 636

Search completed: July 7, 2004, 15:21:06  
Job time : 4.39049 secs

Thu Jul 8 09:12:17 2004

us-09-596-784-4.rpr

---

Page 6



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:07:11 ; Search time 1.96864 Seconds  
(without alignments)  
3676.525 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	81.5	11.4	423	1	PROA_PSEPK
2	80.5	11.3	820	1	AD29 HUMAN
3	79	11.0	789	1	PLSB MYCTU
4	77.5	10.8	775	1	PLSB MYCLE
5	75.5	10.6	966	1	SSN6 YEAST
6	73.5	10.3	320	1	DP44 BRP69
7	73.5	10.3	741	1	DOC8 MOUSE
8	73.5	10.3	976	1	EPAL HUMAN
9	72.5	10.1	421	1	PROA_PSEPM
10	71.5	10.0	1036	1	OGT1 HUMAN
11	71.5	10.0	1036	1	OGT1 RAT
12	71	9.9	816	1	NEL2 MOUSE
13	70.5	9.9	1799	1	DOC8 HUMAN
14	70.5	9.9	2203	1	POLG EC09B
15	70	9.8	946	1	AMPN PLUXY
16	70	9.8	1739	1	DOTL HUMAN
17	69.5	9.7	2183	1	RRPL MEASA
18	69.5	9.7	2183	1	RRPL MEASE
19	69.5	9.7	3680	1	DMD CANFA
20	69	9.7	451	1	YUN HAEN
21	68.5	9.6	173	1	YCF3 SYN3
22	68.5	9.6	586	1	MERL RAT
23	68.5	9.6	595	1	MERL HUMAN
24	68.5	9.6	595	1	MERL PAPAN
25	68.5	9.6	596	1	MERL MOUSE
26	68	9.5	478	1	ARDE CHLMU
27	68	9.5	727	1	DOC7 MOUSE
28	68	9.5	1302	1	DOC7 HUMAN
29	67.5	9.4	404	1	O67C DROME
30	67.5	9.4	1141	1	GSX3 CAEEL
31	67.5	9.4	1527	1	MRP3 HUMAN
32	67	9.4	172	1	MUPL HUMAN
33	67	9.4	532	1	UD14_RABIT

RESULT 1	ID	PROA_PSEPK	STANDARD;	PRT;	423 AA.
AC	Q88DL4;				
DT	15-MAR-2004	(Rel. 43, Created)			
DT	15-MAR-2004	(Rel. 43, Last sequence update)			
DT	15-MAR-2004	(Rel. 43, Last annotation update)			
DE	Gamma-glutamyl phosphate reductase (GPR) (EC 1.2.1.41) (Glutamate-5-semialdehyde dehydrogenase) (Glutamyl-gamma-semialdehyde dehydrogenase) (GSA dehydrogenase).				
GN	PROA OR PP4811.				
OS	Pseudomonas putida (strain KT2440).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OX	Pseudomonadaceae; Pseudomonas.				
NCBI_TaxID=160488;					
SEQUENCE FROM N.A.					
MEDLINE=22423060; PubMed=12534463;					
Nelson K.E., Weinl C., Paulsen I.T., Dodson R.J., Hilbert H., Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M., Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J., Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I., Chris Lee P., Holtzapfel B., Scanlan D., Tran K., Moazzes A., Uterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S., Kiewitz C., Eisen J.A., Timmis J.A., Duesterhoeft A., Tuemmli B., Fraser C.M.;					
"Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440.";					
Environ. Microbiol. 4:799-808(2002).					
CC	FUNCTION: Catalyzes the NADPH dependent reduction of L-gamma-glutamyl 5-phosphate into L-glutamate 5-semialdehyde and phosphate. The product spontaneously undergoes cyclization to form 1-pyrroline-5-carboxylate.				
CC	CATALYTIC ACTIVITY: L-glutamate 5-semialdehyde + phosphate + NADP(+) = L-gamma-glutamyl 5-phosphate + NADPH.				
CC	PATHWAY: Proline biosynthesis; second step.				
CC	SUBCELLULAR LOCATION: Cytoplasmic (by similarity).				
CC	SIMILARITY: Belongs to the gamma-glutamyl phosphate reductase family.				

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; AE016792; AA70380.1; -.  
TIGR; PP4811; -.  
HAMAP; MF\_00412; -; 1.  
InterPro; IPR000965; Gglut\_pp\_reduct.  
PROSITE; PS01223; PROA; 1.  
Oxidoreductase; Proline biosynthesis; NADP; Complete proteome.  
SEQUENCE 423 AA; 45417 MW; E2B9930ABE513A11 CRC64;

```
Query Match 11.4%; Score 81.5; DB 1; Length 423;
Best Local Similarity 20.5%; Pred. No. 1.2;
Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;

QY 12 LQYFSAGCKTPI--HLKDGVCALYNEODEAAVLEVPQHSQ-----50
Db 213 IERISDARVPVFKHL-DGICHY-----VSQHADLDKAWNAPNAKTYRYGIC 260
QY 51 -----SLLHCRITEADPQTSITYLSMLQLNFENWAAWRC-----WLA- 89
Db 261 GAMETLLVDQVAER-----FLPEWARREVEKGVLELRCQRTQAIISAKPATEADWHT 314
QY 90 -LDELHNVRLC--FQSLHDLDEASFSDIIVSGTIEHAAEVREYIAQLDESS 137
Db 315 YLDAILSIRVVDGLNQLAIEHINHYGSHHTDSIISEHQGEARQFPAEVDSSAS 365

RESULT 2
AD29 HUMAN STANDARD; PRT; 820 AA.
AC Q9UKF5; Q9UHP1; Q9UKF3; Q9UKF4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE ADAM 29 precursor (A disintegrin and metalloproteinase domain 29).
GN ADAM29.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS ALPHA; BETA AND GAMMA).
RC TISSUE=Testis;
RX MEDLINE=99443746; PubMed=10512762;
RA Cerretti D.P., DuBoise R.F., Black R.A., Nelson N.;
RT "Isolation of two novel metalloproteinase-disintegrin (ADAM) cDNAs
RT that show testis-specific gene expression."
RL Biochem. Biophys. Res. Commun. 263:810-815(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Testis;
RX MEDLINE=20112771; PubMed=10644455;
RA Xu R., Cai J., Xu T., Zhou W., Ying B., Deng K., Zhao S., Li C.;
RT "Molecular cloning and mapping of a novel ADAM gene (ADAM29) to human
RT chromosome 4."
RL Genomics 62:537-539(1999).
CC -!- FUNCTION: May be involved in spermatogenesis and fertilization.
CC Seems to be a non catalytic metalloprotease-like protein.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=Alpha;
CC IsoId=Q9UKF5-1; Sequence=Displayed;
CC Name=Beta;
CC IsoId=Q9UKF5-2; Sequence=VSP_005491;
CC Name=Gamma;
CC IsoId=Q9UKF5-3; Sequence=VSP_005492, VSP_005493;
CC -!- TISSUE SPECIFICITY: Expressed specifically in testes.
CC -!- SIMILARITY: Contains 1 EGF-like domain.
CC -!- SIMILARITY: Contains 1 disintegrin domain.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF171929; AAF03777.1; -
CC EMBL; AF171930; AAF03778.1; -
CC EMBL; AF171931; AAF03779.1; -
CC EMBL; AF134708; AAF22163.1; -
CC HSSP; P18619; 1FVL.
CC MEROPS; M12.981; -.
CC Genew; HGNC:207; ADAM29.
CC MIM; 604778; -.
CC GO; GO:0005887; C: integral to plasma membrane; TAS.
CC GO; GO:0008237; F: metalloproteinase activity; TAS.
CC GO; GO:0007283; P: spermatogenesis; TAS.
CC GO; GO:0007283; P: spermatogenesis; TAS.
CC InterPro; IPR006586; ADAM_cysteine.
CC InterPro; IPR001762; Disintegrin.
CC InterPro; IPR006025; Pept M Zn BS.
CC InterPro; IPR001590; Peptidase_M12B.
CC Pfam; PF00200; disintegrin; 1.
CC Pfam; PF01562; Pep_M12B_propep; 1.
CC PRINTS; PR00289; Repolysin; 1.
CC ProDom; PD000664; Disintegrin; 1.
CC SMART; SM00050; DISIN; 1.
CC PROSITE; PS50215; ADAM_MEPRO; 1.
CC PROSITE; PS00427; DISINTEGRIN_1; FALSE NEG.
CC PROSITE; PS00427; DISINTEGRIN_2; 1.
CC PROSITE; PS00022; EGF_1; FALSE NEG.
CC PROSITE; PS01186; EGF_2; FALSE NEG.
CC PROSITE; PS00142; ZINC_PROTEASE; FALSE NEG.
CC Signal; Glycoprotein; Transmembrane; EGF-like domain;
KW Alternative splicing; Repeat.
FT SIGNAL 1 18 POTENTIAL.
FT PROPEP 19 193 BY SIMILARITY.
FT CHAIN 194 820 ADAM 29.
FT DOMAIN 194 674 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 675 695 POTENTIAL.
FT DOMAIN 696 820 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 194 390 METALLOPROTEASE-LIKE.
FT DOMAIN 397 483 DISINTEGRIN-LIKE.
FT DOMAIN 484 624 CYS-RICH.
FT DOMAIN 625 654 EGF-LIKE.
FT DOMAIN 659 675 POLY-LYS.
FT DOMAIN 739 819 9 X 9 AA APPROXIMATE REPEATS.
FT REPEAT 739 747 1.
FT REPEAT 748 756 2.
FT REPEAT 757 765 3.
FT REPEAT 766 774 4.
FT REPEAT 775 783 5.
FT REPEAT 784 792 6.
FT REPEAT 793 801 7.
FT REPEAT 802 810 8.
FT REPEAT 811 819 9.
FT DISULFID 307 384 BY SIMILARITY.
FT DISULFID 347 369 BY SIMILARITY.
FT DISULFID 455 468 POTENTIAL.
FT DISULFID 625 636 BY SIMILARITY.
FT DISULFID 630 642 BY SIMILARITY.
FT DISULFID 644 653 BY SIMILARITY.
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 428 428 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 538 538 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 545 545 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 558 558 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 564 564 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 750 803 Missing (in isoform Beta).
FT VARSPLIC 763 787 /FTID=VSP_005491.
FT VARSPLIC 788 820 VMPQSHPQLTSQPPVMPQSHP -> QNLFLEFSISD
CVLNFRLYLOAT (in isoform Gamma).
FT /FTID=VSP_005492.
FT Missing (in isoform Gamma).
FT /FTID=VSP_005493.
```

```

FT CONFLICT 196 H -> Y (IN REF. 2).
FT CONFLICT 744 P -> H (IN REF. 2).
FT CONFLICT 748 S -> Y (IN REF. 2).
FT CONFLICT 753 Q -> R (IN REF. 1); AAF03777.
FT CONFLICT 764 M -> T (IN REF. 2).
FT CONFLICT 769 HPQLT -> QPRVM (IN REF. 2).
SQ SEQUENCE 820 AA; 92753 MW; 1F54E9F8128E4C27 CRC64;

Query Match 11.3%; Score 80.5; DB 1; Length 820;
Best Local Similarity 22.3%; Pred. No. 3.1;
Matches 31; Conservative 28; Mismatches 47; Indels 33; Gaps 6;

QY 18 GCKTPIHLK-----DGVCALYNEQDEAAVLEVPQHSDDLHCRITIEADPQTSITL-- 69
Db 60 GQKHIIHKVKKLSKHLPTTYTDQGAILEDQFVQVNCYYH-GYVEGDPESLSVLS 118
QY 70 ----YSMLQLNFEMAAAMRGWLALDELHNVR-LCFQOQSLHL-----DEASPSDIVSG 118
Db 119 CFGGFOGILQIN-----DFAYEIKPLAFSTFHEHLVYKMDSEKQFSTMRSG 165
QY 119 FIEHAAVREYIAQDESS 137
Db 166 FMQNEITCRMEFEIDNST 184

RESULT 3
PLSB_MYCTU STANDARD; PRT; 789 AA.
ID PLSB_MYCTU AC 053207;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycerol-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR PLSB2 OR RV2482C OR MT2555 OR MTV008.38C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ernolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J.A., Knouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
CC acyl-sn-glycerol 3-phosphate.
CC -!- PATHWAY: De novo phospholipid biosynthesis; first step. May also
CC function in the regulation of membrane biogenesis.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the GPAT / DAPAT family.
CC -!- SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021246; CAAL16059.1; -.
DR EMBL; AE007092; AAK46859.1; -.
DR PIR; A70868; A70868.
DR TIGR; MT2555; -.
DR Tuberculist; RV2482C; -.
DR HAMAP; MF_00393; -.
DR InterPro; IPR002123; Acyltransferase.
DR Pfam; PF01553; Acyltransferase; 1.
DR SMART; SM00563; Plc; 1.
KW Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
KW Complete proteome.
FT CONFLICT 180 R -> Q (IN REF. 2).
FT CONFLICT 778 C -> R (IN REF. 2).
SQ SEQUENCE 789 AA; 88314 MW; B78D7D8F8296EA22 CRC64;

Query Match 11.0%; Score 79; DB 1; Length 789;
Best Local Similarity 24.3%; Pred. No. 4.2;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

QY 23 IHLKDGVCALYNEQDEAAVLEVPQHSDDLHCRITIEADPQTSITLYSMLQLNFEMAA 82
Db 445 VSMRQYLGAPHGELTQDPAKELALQKMSFEVAVRILQATPVTATGLVSALL-----LT 498
QY 83 MRGWLALDELHNVRILCFQOQSLHL-----EASFSDIVSGF----- 119
Db 499 TRGTALTLDQLHHT---LQDSLVDLKRQSPVSTALRLRSRGVRAAADALSNGHPVTR 555
QY 120 IEHAAVREYIAQDESSAA 139
Db 556 VDSGREPVWYIAPDDEHAAA 575

RESULT 4
PLSB_MYCLE STANDARD; PRT; 775 AA.
ID PLSB_MYCLE AC 09X7B0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycerol-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
GN PLSB OR ML1246 OR MLCB1610.07.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
CC acyl-sn-glycerol 3-phosphate.
CC -!- PATHWAY: De novo phospholipid biosynthesis; first step. May also
CC function in the regulation of membrane biogenesis.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the GPAT / DAPAT family.

```

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AL049913; CAB43153.1; -  
 CC DR EMBL; AL583921; CAC31627.1; -  
 CC DR EMBL; T45238; T45238.  
 CC DR Leproma; ML1246; -  
 CC DR HAMAP; MF\_00393; -; 1.  
 CC DR InterPro; IPR002123; Acyltransferase.  
 CC DR Pfam; PF01553; Acyltransferase; 1.  
 CC DR SMART; SM00563; Pfam; 1.  
 CC DR Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;  
 CC KW Complete proteome.  
 CC SQ SEQUENCE 775 AA; 87363 MW; 03DD77C778293CDF CRC64;  
 CC -----  
 CC Query Match 10.8%; Score 77.5; DB 1; Length 775;  
 CC Best Local Similarity 25.8%; Pred. No. 5.9;  
 CC Matches 34; Conservative 16; Mismatches 33; Indels 49; Gaps 5;  
 CC -----  
 CC QY 11 FLOYFSGAGKTPHLKDGVCALYN-----EQ 36  
 CC Db 397 YAYARGGEKTP-----EGVWLKSFKAQGERNYGKIYVFPPEAVSMRQVILGAPHGALVQ 452  
 CC QY 37 DEEAALVLEQPHSDLLHCRIRIBADPQTSITLYSMLLQLNFEMAMRGCGWLALDELHNV 96  
 CC Db 453 DQDAKRLALQKMSPEVA--WRILCATPTATALVSALL-----LITRGVALTLDQLHHT 504  
 CC QY 97 RLCPQOQSLHLD 108  
 CC Db 505 ---LQESLDYLE 513  
 CC -----  
 CC RESULT 5  
 CC SSN6\_YEAST STANDARD; PRT; 966 AA.  
 CC AC P14922;  
 CC DT 01-APR-1990 (Rel. 14, Last sequence update)  
 CC DT 01-APR-1990 (Rel. 14, Last sequence update)  
 CC DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 CC DE Glucose repression mediator protein.  
 CC GN SSN6 OR CYC8 OR YBR112C OR YBR0908.  
 CC OS Saccharomyces cerevisiae (Baker's yeast).  
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CC CX NCBI\_TaxID=4932;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=89211964; PubMed=2854095;  
 CC RA Trumbly R.J.;  
 CC RT "Cloning and characterization of the CYC8 gene mediating glucose  
 CC repression in yeast."  
 CC RL Gene 73:97-111(1988).  
 CC RN [2]  
 CC RP SEQUENCE FROM N.A.  
 CC RX MEDLINE=89065502; PubMed=3316983;  
 CC RA Schultz J., Carlson M.;  
 CC RT "Molecular analysis of SSN6, a gene functionally related to the SNF1  
 CC protein kinase of Saccharomyces cerevisiae."  
 CC RL Mol. Cell. Biol. 7:3637-3645(1987).  
 CC RN [3]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=S288C;  
 CC RX MEDLINE=92327848; PubMed=1626431;  
 CC RA Mannhaupt G., Stucka R., Ehle S., Vetter I., Feldmann H.;  
 CC RT "Molecular analysis of yeast chromosome II between CMD1 and LYS2: the  
 CC RT excision repair gene RAD16 located in this region belongs to a novel  
 CC RT group of double-finger proteins.";

Yeast 8:397-408(1992).  
 [4]  
 RL TPR REPEATS.  
 RN MEDLINE=90124639; PubMed=2404612;  
 RX Sikorski R.S., Boguski M.S., Goebel M., Hieter P.A.;  
 RT "A repeating amino acid motif in CDC23 defines a family of proteins  
 RT and a new relationship among genes required for mitosis and RNA  
 RT synthesis."  
 RL Cell 60:307-317(1990).  
 CC -!- FUNCTION: IT IS INVOLVED IN REPRESSION BY AL-ALPHA2 AND ALPHA2 AND  
 CC IN OTHER SYSTEMS AS A GENERAL REPRESSOR OF TRANSCRIPTION. THIS  
 CC PROTEIN HAS NO OBVIOUS DNA-BINDING DOMAINS. IT MIGHT NOT INTERACT  
 CC DIRECTLY WITH DNA BUT WITH DNA-BOUND PROTEINS.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: Contains 10 TPR repeats.  
 CC -!- SIMILARITY: TO YEAST GAL1 AND CCR4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; M23440; AAA34545.1; -  
 CC DR EMBL; M17826; AAA35103.1; -  
 CC DR EMBL; X66247; CAA46973.1; -  
 CC DR EMBL; X78993; CAA55615.1; -  
 CC DR EMBL; Z35981; CAA85069.1; -  
 CC DR PIR; S25365; S25365.  
 CC DR Germonline; 138655; -  
 CC DR TRANSFAC; T03687; -  
 CC DR SGD; S0000316; CYC8.  
 CC DR GO; GO:0005634; C:nucleus; IPI.  
 CC DR GO; GO:0016565; F:general transcription co-activator activity; IDA.  
 CC DR GO; GO:0003713; F:transcription co-activator activity; IDA.  
 CC DR GO; GO:0016481; P:negative regulation of transcription; IDA.  
 CC DR InterPro; IPR008941; TPR-like.  
 CC DR InterPro; IPR001440; TPR.  
 CC DR Pfam; PF00515; TPR; 10.  
 CC DR SMART; SM00028; TPR; 9.  
 CC DR Transcription regulation; Repressor; Repeat; TPR repeat;  
 CC KW Nuclear protein.  
 CC FT DOMAIN 15 30 POLY-GLN.  
 CC FT REPEAT 46 79  
 CC FT REPEAT 80 113  
 CC FT REPEAT 114 147  
 CC FT REPEAT 150 183  
 CC FT REPEAT 187 220  
 CC FT REPEAT 224 257  
 CC FT REPEAT 258 291  
 CC FT REPEAT 296 329  
 CC FT REPEAT 330 363  
 CC FT REPEAT 364 398  
 CC FT DOMAIN 493 556 30 X 2 AA TANDEM REPEATS OF Q-A.  
 CC FT DOMAIN 557 587 POLY-GLN.  
 CC FT CONFLICT 547 547 K -> Q (IN REF. 3).  
 CC SQ SEQUENCE 966 AA; 107202 MW; 84B509CF3208C5C0 CRC64;  
 CC -----  
 CC Query Match 10.6%; Score 75.5; DB 1; Length 966;  
 CC Best Local Similarity 26.2%; Pred. No. 12;  
 CC Matches 22; Conservative 15; Mismatches 34; Indels 13; Gaps 2;  
 CC -----  
 CC QY 64 QTSITLYSMLLQLNFEM-----AAMRGCGWLALDELHNVRLCPQOQSLHLEAFSDIVSGF 119  
 CC Db 97 QRAAEYERALLVNPFLSDVWATLGHCHYLMDDLQRAYNAVYQQAALYHLSNPNVFKLHWGI 156  
 CC QY 120 -----IFHAAREVREYIAQLD 134  
 CC Db 157 GILYDRYCSLDYAEAEFAKVLIED 180







RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences".  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -!- FUNCTION: Addition of nucleotide-activated sugars directly onto  
CC the polypeptide through O-glycosidic linkage with the hydroxyl of  
CC serine or threonine.  
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + peptide = UDP +  
CC N-acetyl-beta-D-glucosaminyl-peptide.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBUNIT: HETEROTRIMER OF TWO 11 kDa AND ONE 70 kDa SUBUNITS. IT IS  
CC NOT KNOWN IF THE 70 kDa SUBUNIT IS ENCODED BY A SEPARATE GENE OR  
CC IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN  
CC ALTERNATIVE INITIATION OF THE 110 kDa SUBUNIT (BY SIMILARITY).  
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (Possible).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event-Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=O15294-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=O15294-2; Sequence=VSP 006553;  
CC -!- TISSUE SPECIFICITY: Highly expressed in pancreas and to a lesser  
CC extent in skeletal muscle, heart, brain and placenta. Present in  
CC trace amounts in lung and liver.  
CC -!- SIMILARITY: Contains 12 TPR repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----  
CC EMBL; U77413; AAB63466.1; -;  
DR EMBL; AJ315767; CAC86127.1; -;  
DR EMBL; AJ315767; CAC86129.1; -;  
DR EMBL; BC014434; AAH14434.1; -;  
DR Genbank; HGNC:8127; OGT.  
DR MIM; 300255; -;  
DR GO; GO:0005829; C:cytosol; TAS.  
DR GO; GO:0005634; C:nucleus; TAS.  
DR GO; GO:0008375; F:acetylglucosaminyltransferase activity; TAS.  
DR GO; GO:0005515; F:protein binding; TAS.  
DR GO; GO:0006493; P:O-linked glycosylation; TAS.  
DR GO; GO:0007584; P:response to nutrients; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR InterPro; IPR001440; TPR.  
DR SMART; SM00028; TPR; 12.  
DR SMART; SM00028; TPR; 11.  
KW Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;  
KW Alternative splicing  
FT REPEAT 79 112 TPR 1.  
FT REPEAT 113 146 TPR 2.  
FT REPEAT 147 180 TPR 3.  
FT REPEAT 181 214 TPR 4.  
FT REPEAT 215 248 TPR 5.  
FT REPEAT 249 282 TPR 6.  
FT REPEAT 283 316 TPR 7.  
FT REPEAT 317 350 TPR 8.  
FT REPEAT 351 384 TPR 9.  
FT REPEAT 385 418 TPR 10.  
FT REPEAT 419 452 TPR 11.  
FT REPEAT 453 463 TPR 12 (INCOMPLETE).  
FT DOMAIN 477 493 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT VARSPFLIC 1 166 MASSGVNADSTGLAEAHREYQAGDFEAERHCWLMWROE  
PONTGVLLLSLSIHFCRRDKRSAAHFTSLIAIKQNPILAEAY

FT SNLGNVYKRGQOEAIEHYRHALRLKPDIDGIVINLAAL  
FT VAAGDMGQAVAYVSALQYNPDLYCVSRSDLGNNLLKALGRLE  
FT EA -> MLOGHFWLVREGIMISFSPSPNNLFFPLQIFPF  
FT PFTSPSHLSLTPP (in isoform 2).  
FT /FTId=VSP\_006553.  
SQ SEQUENCE 1036 AA; 115705 MW; C3BD67340925A2C2 CRC64;  
Query Match 10.0%; Score 71.5; DB 1; Length 1036;  
Best Local Similarity 24.4%; Pred. No. 32;  
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;  
Qy 24 HLKDGVCALVNEQDEEAALVLEV-----POHSDSL--LHCFRIIADPOTSITLY 70  
Db 281 HFFDAYCNLANALKEGSAEAEVCYNTALRLCTHADSLNNLANIKRQGNIEEARLY 340  
Qy 71 SMILLQNFENAMRGCGWALDELHNVRLCFQSQLEHLEA-----SFSVDVSGFIEHAAE 125  
Db 341 RKALEVEPEFAAHS---NLASVLQOQKQLEALMHYKEAIRISPTFADAYSNMGTLKE 397  
Qy 126 VRE 128  
Db 398 MQD 400  
RESULT 11  
OGTI RAT STANDARD; PRT; 1036 AA.  
AC P56558;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase 110  
kDa subunit (EC 2.4.1.-) (O-GlcNAc transferase p110 subunit).  
GN OGT.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
ON NCBI\_TaxID=10116;  
RX STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=97238869; PubMed=9083067;  
RA Kreppel L.K., Blomberg M.A., Hart G.W.;  
RT "Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and  
RT characterization of a unique O-GlcNAc transferase with multiple  
RT tetratricopeptide repeats."  
RL J. Biol. Chem. 272:9308-9315 (1997).  
CC -!- FUNCTION: Addition of nucleotide-activated sugars directly onto  
CC the polypeptide through O-glycosidic linkage with the hydroxyl of  
CC serine or threonine.  
CC -!- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine + peptide = UDP +  
CC N-acetyl-beta-D-glucosaminyl-peptide.  
CC -!- ENZYME REGULATION: BY TYROSINE PHOSPHORYLATION AND O-GLCNAc  
CC MODIFICATIONS.  
CC -!- PATHWAY: Glycosylation.  
CC -!- SUBUNIT: HETEROTRIMER OF TWO 110 kDa AND ONE 78 kDa SUBUNITS. IT  
CC IS NOT KNOWN IF THE 78 kDa SUBUNIT IS ENCODED BY A SEPARATE GENE  
CC OR IS THE PRODUCT OF EITHER OF A PROTEOLYTIC DEGRADATION OR AN  
CC ALTERNATIVE INITIATION OF THE 110 kDa SUBUNIT.  
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (Possible).  
CC -!- TISSUE SPECIFICITY: Appears to be present in all tissues examined  
CC except kidney.  
CC -!- PTM: AUTOMODIFIED BY O-GLYCOSYLATION WITH O-GLCNAc.  
CC -!- SIMILARITY: Contains 12 TPR repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----



```

DR EMBL; U76557; AAC53121.1; -.
DR PIR; T31673; T31673.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 12.
DR SMART; SM00028; TPR; 11.
KW Transferase; Glycosyltransferase; Nuclear protein; Repeat; TPR repeat;
KW Phosphorylation; Glycoprotein.
FT REPEAT 79 112 TPR 1.
FT REPEAT 113 146 TPR 2.
FT REPEAT 147 180 TPR 3.
FT REPEAT 181 214 TPR 4.
FT REPEAT 215 248 TPR 5.
FT REPEAT 249 282 TPR 6.
FT REPEAT 283 316 TPR 7.
FT REPEAT 317 350 TPR 8.
FT REPEAT 351 384 TPR 9.
FT REPEAT 385 418 TPR 10.
FT REPEAT 419 452 TPR 11.
FT REPEAT 453 483 TPR 12 (INCOMPLETE).
FT DOMAIN 478 493 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 979 979 PHOSPHORYLATION (PROBABLE).
SQ SEQUENCE 1036 AA; 115605 MW; 3F057CABDD019BD6 CRC64;

Query Match
Best Local Similarity 10.0%; Score 71.5; DB 1; Length 1036;
Matches 30; Conservative 21; Mismatches 51; Indels 21; Gaps 4;

QY 24 HLKDGVCALYNQDEEAAVLEV-----POHSDSL--LLHCRILEADPQTSITLY 70
Db 281 HFPDAYCNLANALKKGVAEEDCVNTRALRCLPTHADSLNNLANIKRQGNIEAAVLY 340

QY 71 SMLQLQNFEMAAMRCGLWALDELHNVRLCFQOSLEHLDEA----SFSDIVSGFIEHAAE 125
Db 341 RKALEVFEFAAHS---NLASVLQOQGLQEAALMHYKSAIRISPTFADAYSNMGNLTKE 397

QY 126 VRE 128
Db 398 MQD 400

RESULT 12
NEL2 MOUSE
ID NEL2_MOUSE STANDARD; PRT; 816 AA.
AC Q61220;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein kinase C-binding protein NELL2 precursor (NEL-like protein 2)
DE (MEL91 protein).
GN NELL2 OR MEL91.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1] _TaxID=10090;
SEQUENCE FROM N.A.
RA Elkins D.A., Rossi J.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBSJ databases.
RP [2]
RP TSP N-TERMINAL DOMAIN.
RX MEDLINE=98153258; PubMed=9480764;
RA Beckmann G., Hanke J., Bork P., Reich J.;
RA "Merging extracellular domains: fold prediction for laminin G-like
RT and amino-terminal thrombospondin-like modules based on homology to
RT pentraxins.";
RL J. Mol. Biol. 275:725-730(1998).
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC -!- SIMILARITY: Contains 5 VWFC domains.
CC -!- SIMILARITY: Contains 6 EGF-like domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC with the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC EMBL; U59230; AAB02924.1; ALT_INIT.
DR HSSP; P00740; IEDM.
DR MGD; MGI:1858510; Nell2.
DR InterPro; IPR000152; Asx_hydroxyl_S.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF00008; EGF; 4.
DR Pfam; PF02210; TSPN; 1.
DR Pfam; PF00093; VWFC; 2.
DR SMART; SM00179; EGF_Ca; 3.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR SMART; SM00214; VWFC; 2.
DR PROSITE; PS00010; ASX HYDROXYL; 3.
DR PROSITE; PS00022; EGF 1; 1.
DR PROSITE; PS01186; EGF 2; 3.
DR PROSITE; PS00026; EGF 3; 6.
DR PROSITE; PS01187; EGF_Ca; 3.
DR PROSITE; PS01208; VWFC_1; 2.
DR PROSITE; PS01184; VWFC_2; 2.
KW Glycoprotein; EGF-like domain; Repeat; Signal.
FT SIGNAL 1 24
FT CHAIN 25 816
FT DOMAIN 30 258
FT DOMAIN 272 331
FT DOMAIN 332 396
FT DOMAIN 397 439
FT DOMAIN 440 481
FT DOMAIN 482 522
FT DOMAIN 521 553
FT DOMAIN 555 601
FT DOMAIN 602 637
FT DOMAIN 638 693
FT DOMAIN 698 756
FT DOMAIN 758 813
FT DISULFID 401 413
FT DISULFID 407 422
FT DISULFID 424 438
FT DISULFID 444 457
FT DISULFID 451 466
FT DISULFID 468 480
FT DISULFID 486 499
FT DISULFID 493 508
FT DISULFID 510 521
FT DISULFID 525 535
FT DISULFID 529 541
FT DISULFID 543 552
FT DISULFID 559 572
FT DISULFID 566 581
FT DISULFID 583 600
FT DISULFID 606 619
FT DISULFID 613 628
FT DISULFID 630 636
FT CARBOHYD 53 53
FT CARBOHYD 225 225
FT CARBOHYD 293 293
FT CARBOHYD 298 298
FT CARBOHYD 517 517
FT CARBOHYD 615 615
FT CARBOHYD 635 635
SQ SEQUENCE 816 AA; 91163 MW; 5BDD0A946F87E74D CRC64;

9.9%; Score 71; DB 1; Length 816;
Query Match

```

Best Local Similarity 24.6%; Pred. No. 28;  
Matches 33; Conservative 12; Mismatches 55; Indels 34; Gaps 4;

QY 3 SSQORVERFLQ-----YFAGCKPIHLKDGVCALYNQDEEAAVLEVPQHSDSLH 55  
Db 69 ASATATATFLQKLNKHEFTILVTLKQHLNLSGVILSIHLDHRYLESSEGRNEIRLH 128

QY 56 CRRIEADPQTSITLYSMLLQNFEMAAAMRGWCWALDELHNVRLCFQOS--LEHLD----- 108  
Db 129 YRSGTHRPHTVEVPYI-----LADAKWHKLSAFSASHLILHIDCKNIY 172

QY 109 -----EASFSDIVSG 118  
Db 173 ERVVEPFTDLALG 186

RESULT 13  
DOC8 HUMAN STANDARD; PRT; 1799 AA.  
AC QNF50; Q8TEP1; Q8WUY2; Q8BYU5; Q9HUQ2; Q9H1Q3; Q9H308; Q9H7P2;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Dedicator of cytokinesis protein 8 (fragment).  
GN DCK8.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RN SEQUENCE FROM N.A.  
RP TISSUE=Spleen;  
RA Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;  
RT "The nucleotide sequence of a long cDNA clone isolated from human  
spleen.";  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
[2]  
RN SEQUENCE FROM N.A.  
RA Laird G., Skuce C.;  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RN SEQUENCE OF 1194-1799 FROM N.A.  
RP TISSUE=Muscle;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,  
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
Raba S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,  
Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,  
Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[4]  
RN SEQUENCE OF 1475-1799 FROM N.A.  
RX MEDLINE=20195627; PubMed=10729223;  
RA Oltenghi C., Veitia R., Quintana-Murci L., Torchard D., Scapoli L.,  
Souleyreau-Therville N., Beckmann J., Fellous M., McElreavey K.;  
RT "The region on 9p associated with 46,XY sex reversal contains several  
transcripts expressed in the urogenital system and a novel  
doublesex-related domain.";  
RL Genomics 64:170-178 (2000).

RN SEQUENCE OF 1595-1799 FROM N.A.  
 RP TISSUE=Lymph node;  
 RA Koehrer K., Beyer A., Mewes H.-W., Weil B., Wiemann S.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 [6]  
 RN NOMENCLATURE.  
 RX MEDLINE=22319137; PubMed=12432077;  
 RA Cote J.-F., Vuori K.;  
 RT "Identification of an evolutionarily conserved superfamily of DCK180-  
 related proteins with guanine nucleotide exchange activity.";  
 J. Cell Sci. 115:4901-4913 (2002).  
 CC -1- FUNCTION: Potential guanine nucleotide exchange factor (GEF). GEF  
 proteins activate some small GTPases by exchanging bound GDP for  
 free GTP (By similarity).  
 CC -1- DOMAIN: The DHR-2 domain may mediate some GEF activity (By  
 similarity).  
 CC -1- SIMILARITY: Belongs to the DCK family.  
 CC -1- SIMILARITY: Contains 1 DHR-1 (CZH-1) domain.  
 CC -1- SIMILARITY: Contains 1 DHR-2 (CZH-2) domain.  
 CC -1- CAUTION: Ref.1 (BAB84907) sequence differs from that shown due to  
 a frameshift in position 1656.  
 CC -1- CAUTION: Ref.2 (CAC22149) sequence differs from that shown due to  
 erroneous gene model prediction.  
 CC -1- CAUTION: Ref.4 sequence differs from that shown due to  
 frameshifts in positions 1767 and 1782.  
 ---  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 ---  
 DR EMBL; AK024436; BAB15726.1; -  
 DR EMBL; AK074081; BAB84907.1; ALT\_FRAME.  
 DR EMBL; AK090429; BAC03410.1; -  
 DR EMBL; AL158832; -; NOT ANNOTATED\_CDS.  
 DR EMBL; AL161725; CAC22148.1; -  
 DR EMBL; AL161725; CAC22149.1; ALT\_SEQ.  
 DR EMBL; BC019102; AAH19102.1; ALT\_INIT.  
 DR EMBL; AF194407; AAC42221.1; ALT\_FRAME.  
 DR EMBL; AL583913; CAC29497.1; -  
 DR EMBL; HGNC:19191; DCK8.  
 KW Guanine-nucleotide releasing factor.  
 FT NON\_TER 1  
 FT DOMAIN 260 502 DHR-1.  
 FT DOMAIN 1234 1755 DHR-2.  
 FT CONFLICT 568 569 SG -> C (IN REF. 2).  
 FT CONFLICT 1451 1451 V -> F (IN REF. 1; BAB84907).  
 FT CONFLICT 1455 1455 V -> F (IN REF. 1; BAB84907).  
 FT CONFLICT 1670 1670 A -> P (IN REF. 4).  
 FT CONFLICT 1729 1729 E -> K (IN REF. 4).  
 FT CONFLICT 1746 1746 L -> F (IN REF. 4).  
 SQ SEQUENCE 1799 AA; 204390 MW; 29FECA062A0E16 CRC64;  
 Query Match 9.9%; Score 70.5; DB 1; Length 1799;  
 Best Local Similarity 21.6%; Pred. No. 77;  
 Matches 29; Conservative 24; Mismatches 42; Indels 39; Gaps 6;  
 QY 24 HKDGVCMAL--YFQDEEAAVLEVPQHSDSL--LHCRIEADPQTSITLYS 71  
 Db 1269 HLRSRLTILAYSEDTAMQTPPTQVEELCNLSILYDTVKMREFQDPE----- 1321  
 QY 72 MLLQLNPEMAA-----MRGCWLA-LDELHNVRLCFQOSLEHLDSEAFSDIVSGFIEHA 123  
 Db 1322 MMDLMYRIAKSYQASPDRLRLTLQNLQNAEKKTKKCYTEAMCL-----VHAA 1369  
 QY 124 AEVREYIAQLDESS 137  
 Db 1370 ALVAEYLSMLEDHS 1383

RESULT 14  
POLG\_EC09B  
ID POLG\_EC09B STANDARD; PRT; 2203 AA.  
AC Q66577;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein (Contains: Coat protein VP4 (PIA); Coat protein VP2 (PIB); Coat protein VP3 (PIC); Coat protein VP1 (PID); Picornain 2A (EC 3.4.22.29) (P2A); Core protein P2B; Core protein P2C; Core protein P3A; Genome-linked protein VPg (P3B); Picornain 3C (EC 3.4.22.28) (Protease 3C) (P3C); RNA-directed RNA polymerase (EC 2.7.7.48) (P3D)].  
DE Echovirus 9 (strain Barty).  
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Picornaviridae;  
OC Enterovirus.  
OX NCBI\_TaxID=103914;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97033384; PubMed=8879131;  
RA Zimmermann H., Eggers H.J., Nelsen-Salz B.;  
RT "Molecular cloning and sequence determination of the complete genome of the virulent echovirus 9 strain Barty.";  
RT Virus Genes 12:149-154(1996).  
RL  
CC -!- FUNCTION: P2A AND THE P3C POLYPEPTIDES ARE PROTEASES THAT CLEAVE AT CERTAIN Q/G SITES IN THE POLYPROTEIN. THEY ARE CYSTEINE PROTEASES.  
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Gln-|-Gly bond in the poliovirus polyprotein. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.  
CC -!- CATALYTIC ACTIVITY: Selective cleavage of Tyr-|-Gly bond in the picornavirus polyprotein. In other picornavirus reactions Glu may be substituted for Gln, and Ser or Thr for Gly.  
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + [RNA] [N].  
CC -!- SUBUNIT: The virus capsid is composed of 60 icosahedral units, each of which is composed of one copy each of proteins VP1, VP2, VP3, and VP4.  
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins. CLEAVAGE BETWEEN VP4 AND VP2 IS AUTOCATALYTIC; VP1/P2A IS CATALYZED BY P2A; ALL OTHER CLEAVAGES ARE CATALYZED BY P3C.  
CC -!- SIMILARITY: P2A PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC -!- SIMILARITY: P3C PROTEASE BELONGS TO PEPTIDASE FAMILY C3.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X92886; CAA63480.1; -.  
CC HSPF; P21404; IDAM.  
CC MEROPS; C03.001; -.  
CC InterPro; IPR003593; AAA ATPase.  
CC InterPro; IPR004004; Calici pol hel.  
CC InterPro; IPR009003; Cys Ser trypsin.  
CC InterPro; IPR000199; Pept\_3C\_picorn.  
CC InterPro; IPR000081; Peptidase\_C3.  
CC InterPro; IPR003138; Pico\_P1A.  
CC InterPro; IPR002527; Pico\_P2B.  
CC InterPro; IPR001676; Rhv.  
CC InterPro; IPR000605; RNA helicase.  
CC InterPro; IPR007095; RNA\_pol\_DS\_PS.  
CC InterPro; IPR001205; RNA\_pol\_P3D.  
CC InterPro; IPR007094; RNA\_pol\_PSVir.  
CC InterPro; IPR008975; Viral\_cap\_coat.  
DR Pfam; PF00548; Cys-protease-3C; 1.  
DR Pfam; PF02226; Pico\_P1A; 1.  
DR Pfam; PF00947; Pico\_P2A; 1.  
DR Pfam; PF01552; Pico\_P2B; 1.  
DR Pfam; PF00073; rhv; 3.

DR Pfam; PF00680; RNA\_dep\_RNA\_pol; 1.  
DR Pfam; PF00910; RNA\_helicase; 1.  
DR PRINTS; PR00918; CALICVIRUSNS.  
DR ProDom; PD001125; Cys\_protease\_3C; 1.  
DR ProDom; PD001306; Pico\_P2A; 1.  
DR ProDom; PD001274; Pico\_P2B; 1.  
DR SMART; SM00382; AAA; 1.  
KW Polyprotein; Coat protein; Core protein; Transferase; Myristate;  
KW RNA-directed RNA polymerase; Hydrolase; Thiol protease; Lipoprotein.  
FT CHAIN 2 69  
FT CHAIN 70 330  
FT CHAIN 331 569  
FT CHAIN 570 869  
FT CHAIN 870 1019  
FT CHAIN 1020 1118  
FT CHAIN 1119 1447  
FT CHAIN 1448 1536  
FT CHAIN 1537 1558  
FT CHAIN 1559 1741  
FT CHAIN 1742 2203  
FT LIPID 2 2  
FT ACT\_SITE 1705 1705  
FT ACT\_SITE 1719 1719  
FT ACT\_SITE 1719 1719  
SQ SEQUENCE 2203 AA; 246279 MW; 79B78F8C99D02777 CRC64;  
  
Query Match 9.9%; Score 70.5; DB 1; Length 2203;  
Best Local Similarity 24.3%; Pred. No. 98;  
Matches 33; Conservative 17; Mismatches 57; Indels 29; Gaps 5;  
  
QY 12 LQYFSAGCKTPIHLKD---GVCALYNEODEEAAVLEVPQHSDDLHLHCRIIEADPQTSIT 68  
DB 463 LAYSHAGASVPSRRDAMLGTHVIWDVGLQSCVLCVFWISQT---HYRLVAQDEYTS-- 517  
QY 69 LYSMLQLNPFMAAMRGCGWLDEL-----HNVRLCFQOSLEHLDSEASDVIUSGPIE 121  
DB 518 -----AGYITCQYTNIVVPETPESDCVLCVFSACNDFSVRLKD--TPPIE 563  
QY 122 HAAEVREYIAQLDESS 137  
DB 564 QAELQNDVROAVEGA 579  
  
RESULT 15  
AMPN\_PLUXY  
ID AMPN\_PLUXY STANDARD; PRT; 946 AA.  
AC P91887;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Aminopeptidase N precursor (EC 3.4.11.2) (Microsomal aminopeptidase (APN1)).  
DE APN1.  
GN APN1.  
OS Plutella xylostella (Diamondback moth).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
OC Yponomeutoidea; Plutellidae; Plutella.  
OX NCBI\_TaxID=51655;  
RN [1]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC TISSUE=Midgut;  
RX MEDLINE=98000273; PubMed=9342226;  
RA Denolf P.H., Hendrickx K., van Damme J., Jansens S., Peferoen M.,  
RA Degheele D., van Rie J.;  
RT "Cloning and characterization of Manduca sexta and Plutella xylostella midgut aminopeptidase N related to Bacillus thuringiensis toxin-binding proteins.";  
RL Eur. J. Biochem. 248:748-761(1997).  
CC -!- CATALYTIC ACTIVITY: Release of an N-terminal amino acid, Xaa-|-Xbb- from a peptide, amide or arylamide. Xaa is preferably Ala, but may be most amino acids including Pro (slow action). When a terminal hydrophobic residue is followed by a prolyl residue, the two may be released as an intact Xaa-Pro dipeptide.

CC -|- COFACTOR: Binds 1 zinc ion per subunit (By similarity).  
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.  
CC -|- SIMILARITY: Belongs to peptidase family M1.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
CC EMBL; X97878; CAA66467.1; --  
CC MEROPS; M01.013; --  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR InterPro; IPR001930; Peptidase\_M1.  
DR Pfam; PF01433; Peptidase\_M1; 1.  
DR PRINTS; PR00756; ALADIPTASE.  
DR PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc; Glycoprotein;  
KW GPI-anchor; Signal.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 ? AMINOPEPTIDASE N.  
FT PROPEP ? 946 REMOVED IN MATURE FORM (POTENTIAL).  
FT METAL 344 344 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT SITE 345 345 BY SIMILARITY.  
FT METAL 348 348 ZINC (CATALYTIC) (BY SIMILARITY).  
FT METAL 367 367 ZINC (CATALYTIC) (BY SIMILARITY).  
FT ACT SITE 429 429 PROTON DONOR (POTENTIAL).  
FT CARBOHYD 60 60 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 550 550 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 605 605 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 946 AA; 106577 MW; 1DFD81A364067BFB CRC64;

Query Match 9.8%; Score 70; DB 1; Length 946;  
Best Local Similarity 20.6%; Pred. No. 41;  
Matches 33; Conservative 25; Mismatches 62; Indels 40; Gaps 5;

QY 19 CKTPIHLKGVCAVNEQDE-----EAAVLEVPQHSDSL--LLHCRILEADPQ 64  
Db 14 CGNPVQLTDSIALQNTYDNYVLPGESFPTFYDQVLFDFEYASFGTVAIRVVPRIAT 73  
QY 65 TSITLYSMLIQL-----NPEMAAMRGWLALDELHNVRILCFQOSLEHLD--- 108  
Db 74 QEIVLHAMEMELISIRAYSOLPSDDNLNENLFSSYTLATDDTHLLKIQFTRVLDALQPIIT 133  
QY 109 -EASFSD-----IVSGFIEHAAEVREYIAQLDESSA 138  
Db 134 VEISYSAQYAPNMFVGVVYSRVENGATVSLVTSOLOPTFA 173

Search completed: July 7, 2004, 15:18:39  
Job time : 3.96864 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 7, 2004, 15:13:22 ; Search time 6.18715 Seconds  
(without alignments)  
7088.403 Million cell updates/sec

Title: US-09-596-784-4

Perfect score: 715

Sequence: 1 MTSSQQRVERFLQYFSAGCK.....IEHAAEVREYIAQLDESSAA 139

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organalle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_todent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	715	100.0	139	2	054621 erwinia amy
2	455	63.6	124	2	09KH43
3	442	61.8	139	2	09FCV6
4	273.5	38.3	129	2	066102
5	273.5	38.3	131	16	Q9JF37
6	86.5	12.1	142	16	Q88BP9
7	81.5	11.4	423	16	Q88DL4
8	81	11.3	186	16	Q92RQ7
9	81	11.3	723	17	Q8ZW74
10	80.5	11.3	368	10	Q40879
11	79.5	11.1	1021	16	Q8D7R9
12	79	11.0	708	16	Q81R76
13	79	11.0	789	16	Q7TVH5
14	78.5	11.0	830	16	Q9P332
15	78	10.9	155	2	Q8RP06
16	78	10.9	830	16	Q87E45

17	77.5	10.8	300	16	Q8F9X7
18	77	10.8	130	16	Q87W45
19	76.5	10.7	748	4	Q86VZ7
20	76	10.6	371	10	Q8RVMO
21	75.5	10.6	212	16	Q8A7K9
22	75.5	10.6	402	5	O45260
23	75.5	10.6	468	16	Q8PN02
24	75.5	10.6	1588	11	Q9ESK9
25	75	10.5	317	2	Q9FD53
26	75	10.5	317	16	Q97S10
27	75	10.5	923	13	Q7ZVMO
28	75	10.5	2077	5	Q9XYM2
29	74.5	10.4	549	5	Q86SC9
30	74.5	10.4	1051	5	Q9VA99
31	74.5	10.4	1109	5	Q8IM19
32	74	10.3	304	16	Q8PBE7
33	74	10.3	443	2	Q93A56
34	73.5	10.3	751	11	Q7TMQ5
35	73.5	10.3	892	4	Q9HCF0
36	73	10.2	578	11	Q62693
37	73	10.2	643	10	Q9FTE5
38	73	10.2	678	16	Q87MS2
39	73	10.2	1708	16	Q8YZI2
40	72.5	10.1	131	16	Q8YVY8
41	72.5	10.1	421	16	Q87VV6
42	72.5	10.1	678	11	Q8BRY9
43	72.5	10.1	1467	11	Q8CHH8
44	72.5	10.1	1958	2	Q93CG6
45	72.5	10.1	2183	12	Q98588

#### ALIGNMENTS

#### RESULT 1

ID	O54621	PRELIMINARY;	PRT;	139 AA.
AC	O54621;			
DT	01-JUN-1998	(TRENBLrel. 06, Created)		
DT	01-JUN-1998	(TRENBLrel. 06, Last sequence update)		
DT	01-JUN-2003	(TRENBLrel. 24, Last annotation update)		
DE	DSPP.			
GN	DSPP OR DSPP.			
OS	Erwinia amylovora.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Erwinia.			
OX	NCBI_TaxID=552;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EA321;			
RX	MEDLINE=98115919; PubMed=9448330;			
RA	Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,			
RA	Conlin A.K., Collmer A., Beer S.V.;			
RT	"Homology and functional similarity of an hrp-linked pathogenicity			
RT	locus, dsrEF, of Erwinia amylovora and the avirulence locus avrE of			
RT	Pseudomonas syringae pathovar tomato."			
RL	Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EA321;			
RA	Bogdanove A.J., Kim J.F., Wei Z.-M., Kolchinsky P., Beer S.V.;			
RA	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CFBP1430;			
RX	MEDLINE=98086111; PubMed=9426142;			
RA	Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;			
RT	"DsrA, an essential pathogenicity factor of Erwinia amylovora showing			
RT	homology with AvrE of Pseudomonas syringae, is secreted via the Hrp			
RT	secretion pathway in a DsrB-dependent way."			
RL	Mol. Microbiol. 26:1057-1069(1997).			
DR	EMBL; U97504; AAC04851.1; -.			
DR	EMBL; Y13831; CAA74157.1; -.			

Q8f9x7 leptospira  
Q87w45 pseudomonas  
Q86vz7 homo sapien  
Q8rvmo coffea cane  
Q8a7k9 bacteroides  
O45260 caenorhabdi  
Q8pn02 xanthomonas  
Q9esk9 mus musculu  
Q9fd53 streptococc  
Q97ei0 streptococc  
Q7zvm0 brachydanio  
Q9xym2 drosophila  
Q86sc9 ciona intes  
Q9va99 drosophila  
Q8im19 drosophila  
Q8pbe7 xanthomonas  
Q93a56 pseudomonas  
Q7lmq5 mus musculu  
Q9hcf0 homo sapien  
Q62693 rattus norv  
Q9fte5 oryza sativ  
Q87ms2 vibrio para  
Q8yz12 anabaena sp  
Q87vv6 pseudomonas  
Q8bry9 mus musculu  
Q8chh8 mus musculu  
Q93cg6 photobacter  
Q98588 subacute sc





DR GO: 0004350; F-glutamate-5-semialdehyde dehydrogenase acti. . . ; IEA.  
DR GO: 0005561; P-proline biosynthesis; IEA.  
DR InterPro: IPR000965; Gblut\_pp\_reduct.  
DR PROSITE: PS01223; PROA; 1.  
KW Complete proteome.  
SQ SEQUENCE 423 AA; 45417 MW; E2B9930ABE513A11 CRC64;

Query Match 11.4%; Score 81.5; DB 16; Length 423;  
Best Local Similarity 20.5%; Pred. No. 2.9; Indels 63; Gaps 8;  
Matches 35; Conservative 28; Mismatches 45; Indels 63; Gaps 8;

QY 12 LQVFSAGCKTPI--HLKDGVCALYNEQDEAAVLEVPQHSQ-----VSQADLLKANNVAFNAKTYRYGIC 260  
DB 213 IERISDARVPIKHL-DGICHYI-----VSQADLLKANNVAFNAKTYRYGIC 260

QY 51 ----SILLHCRITFADPQTSITLYSMLLQNFEMAMRGCG-----WLA- 89  
DB 261 GAMETLLVDQVAER-----FLPEMARFRVEKGVELRGCGTCQAISAKPATEADWHITE 314

QY 90 -LDELHNVLRC--FQSLHLEDAFSDIVSGFIEHAAEVREYIAQLDES 137  
DB 315 YLDAILLSIRVQGLNQAIHINHYGSHHTDSIISEHQGEARQFMAEVDSSAS 365

RESULT 8  
Q92RQ7 PRELIMINARY; PRT; 186 AA.

ID Q92RQ7  
AC Q92RQ7  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein R00802.  
GN R00802 OR SMC00902.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,  
RA Pohl T., Potetelle D., Puchler A., Purnelle B., Rampeger U.,  
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont  
RT Sinorhizobium meliloti strain 1021";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).  
DR EMBL: AL591785; CAC45374.1; --  
DR GO: 0003677; P:DNA binding; IEA.  
DR GO: 0008170; F:N-methyltransferase activity; IEA.  
DR GO: 0006306; P:DNA methylation; IEA.  
DR InterPro: IPR004398; Cons.hypoth95.  
DR InterPro: IPR002052; N6 Mtase.  
DR Pfam: PF03602; Cons.hypoth95; 1.  
DR TIGRFAMs: TIGR00095; TIGR00095; 1.  
DR PROSITE: PS00092; N6\_MTASE; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 186 AA; 20282 MW; 19B53236451C6EB6 CRC64;

Query Match 11.3%; Score 81; DB 16; Length 186;  
Best Local Similarity 25.0%; Pred. No. 1.3;  
Matches 34; Conservative 19; Mismatches 41; Indels 42; Gaps 5;

QY 12 LQVFSAGCKTPIHLKDGVCALYNEQDEAAVLEVPQHSQSLHCR----- 57  
DB 61 LEALSRCGRQVLFVEQGV-----EGRLIRI--NIEALGLQGRKIFRRDATDLGFV 110

QY 58 -----IIEADPQTSITLYSMLLQNFEMAMRGCMIA-----LDELHNVLRCFQOQL 104  
DB 111 GTMEFFHLVADFPYKGLGERALS-----AARGGMLVPGALALIEADRVRFQFSF 165

QY 105 EHLDEAFSDIVSGFI 120  
DB 166 ESDERAFGDTLMHFL 181

RESULT 9  
Q8ZW74 PRELIMINARY; PRT; 723 AA.

ID Q8ZW74  
AC Q8ZW74  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Carbon monoxide dehydrogenase large subunit, conjectural.  
GN PAE1935.  
OS Pyrobaculum aerophilum.  
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
OC Thermoproteaceae; Pyrobaculum.  
OX NCBI\_TaxID=13773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IM2 / ATCC 51768 / DSM 7523;  
RX MEDLINE=21664397; PubMed=11792869;  
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
RA Miller J.H.;  
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
RT aerophilum";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989 (2002).  
DR EMBL: AF009848; AAL63828.1; --  
DR GO: 0016491; F:oxidoreductase activity; IEA.  
DR GO: 0006118; P:electron transport; IEA.  
DR InterPro: IPR008274; Aldxan dh bind.  
DR InterPro: IPR000674; Aldxan dh hamm.  
DR Pfam: PF01315; Ald\_Xan dh C; 1.  
DR Pfam: PF02738; Ald\_Xan dh\_C2; 1.  
KW Complete proteome.  
SQ SEQUENCE 723 AA; 78914 MW; FCA65FADDE9FB235 CRC64;

Query Match 11.3%; Score 81; DB 17; Length 723;  
Best Local Similarity 24.6%; Pred. No. 6.1;  
Matches 34; Conservative 27; Mismatches 53; Indels 24; Gaps 6;

QY 15 FSGACKTPIHLKDGVCALYNEQDEAAVLEVPQHSQSLHCRITFADPQTSITLYS 71  
DB 537 FSCGKGSAT-VKDVVRVVRGEAAQTVEAIYHADSTFFGVHIAVLELDPEIGF-VKP 594

QY 72 MLIQLNFEMAMRGCMIALDELHNVLRCFQOQLHLEDAFSDIV---SG----- 118  
DB 595 MLKSYDDVGVVNPPLASQITG-----GALQGIQAALYEEVYDESGNLTSLNLA 648

QY 119 FIEHAAEVREYIAQLDES 136  
DB 649 YVPTAAEPKVEYVFAES 666

RESULT 10  
Q40879 PRELIMINARY; PRT; 368 AA.

ID Q40879  
AC Q40879  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Myristyl-ACP desaturase.  
OS Pelargonium hortorum (Common geranium).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC Geraniales; Geraniaceae; Pelargonium.  
OX NCBI\_TaxID=4031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=line: 88-51-10; TISSUE=Trichome gland;  
RX MEDLINE=96323300; PubMed=8710947;  
RA Schultz D.J., Cahoon E.B., Shanklin J., Craig R., Cox-Foster D.L.,  
RA Mumma R.O., Medford J.I.;





```

RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL: BX248342; CAD97368.1; -.
KW Transferase; Acyltransferase; Complete proteome.
SQ SEQUENCE 789 AA; 88340 MW; 85ACA549E9B1CFPI CRC64;

Query Match 11.0%; Score 79; DB 16; Length 789;
Best Local Similarity 24.3%; Pred. No. 11;
Matches 34; Conservative 23; Mismatches 51; Indels 32; Gaps 4;

QY 23 IHKQGVCAALYNEQDEAAVLEVPQHSDDLHLCRIIEADPQTSITLYSMLLQNFEMAA 82
DB 445 VSMRQVLGAPGHELTQDPAKRLALQKMSPEVAMRIQLQATPVATGLVSALL-----LT 498
QY 83 MRCQWLALDELHNVRLCFQOSLEHLD-----EASFSDIVSGF----- 119
DB 499 TRGVATLDQGHHT---LQSDLDYLERKQSPVSTSLRLRSREGVRAAALSGNHPVTR 555
QY 120 IEHAAEVREYIAQLDESSAA 139
DB 556 VDSGREFVWIAPDDEHAAA 575

RESULT 14
Q9PE32 PRELIMINARY; PRT; 830 AA.
AC Q9PE32;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE Ribonucleoside-diphosphate reductase large chain (EC 1.17.4.1).
GN XF1196.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9a5C;
MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lenos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.F.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

```

```

RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
CC -1- FUNCTION: PROVIDES THE PRECURSORS NECESSARY FOR DNA SYNTHESIS (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: 2'-DEOXYRIBONUCLEOSIDE DIPHOSPHATE + OXIDIZED
CC THIOREDOXIN + H(2)O = RIBONUCLEOSIDE DIPHOSPHATE + REDUCED
CC THIOREDOXIN.
CC -1- PATHWAY: FIRST REACTION IN THE DNA REPLICATION PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEOSIDE DIPHOSPHATE REDUCTASE
CC LARGE CHAIN FAMILY.
DR EMBL: AE003954; AAF84006.1; -.
DR PIR: C82710; C82710.
DR GO: GO:0005971; C:ribonucleoside-diphosphate reductase complex; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0004748; F:ribonucleoside-diphosphate reductase activity; IEA.
DR GO: GO:0006260; P:DNA replication; IEA.
DR InterPro: IPR000788; Ribonucleo_red.
DR InterPro: IPR008926; Ribonucleo_red_N.
DR Pfam: PF00317; ribonuc_red_lg; 1.
DR Pfam: PF02867; ribonuc_red_lgc; 1.
DR PRINTS: PR01183; RIBORDTASEM1.
DR PROSITE: PS00089; RIBORED_LARGE; 1.
KW DNA replication; Oxidoreductase; Complete proteome.
SQ SEQUENCE 830 AA; 94229 MW; 12207407246E28BC CRC64;

Query Match 11.0%; Score 78.5; DB 16; Length 830;
Best Local Similarity 25.6%; Pred. No. 13;
Matches 33; Conservative 21; Mismatches 46; Indels 29; Gaps 6;

QY 17 AGCKTPHLKDGVCALYNEQDEAAVLEVPQHSDDLHLCRII-EADPQTSITLYSMLLQ 75
DB 22 AGSVPPVALS-----VFHPDEPD-----EVPIPRQATMMHTRAVEANVSTWITACGNRR 73
QY 76 LNFEMAAMRCGLALDELHNVRLCFQOSLEHLDSEAFSDIVSGFIEHAAEV----- 126
DB 74 MPFDCARLE---CAIDTIH-----QEFQLDVAEYKRAVGFVEVERKDSVNADDIVDLL 123
QY 127 -REYIAQLD 134
DB 124 IREAEARVD 132

RESULT 15
Q8RP06 PRELIMINARY; PRT; 155 AA.
AC Q8RP06;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Putative chaperone ChpPmak.
OS Pseudomonas syringae pv. maculicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=59511;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ES4326;
MEDLINE=21862332; PubMed=11872842;
RA Guttman D.S., Vinatzer B.A., Sarkar S.F., Ranall M.V., Kettler G.,
RA Greenberg J.F.;
RT "A Functional Screen for the Type III (Hrp) Secretome of the Plant
RT Pathogen Pseudomonas syringae.";
RL Science 295:1722-1726(2002).
DR EMBL: AF458049; AAL84250.1; -.
SQ SEQUENCE 155 AA; 17124 MW; 33EF90778D00E809 CRC64;

Query Match 10.9%; Score 78; DB 2; Length 155;
Best Local Similarity 26.3%; Pred. No. 2.1;
Matches 31; Conservative 21; Mismatches 48; Indels 18; Gaps 5;

QY 17 AGCKTPHLKDGVCALYNEQDEAAVLEVPQHSDDLHLCRII-EADPQTSITLYSMLLQ 76
DB 24 AGCVD-----QGHSA-FSDREGNLFQAVDPADFSSLHLYLTLELQAVDGRVLASALAI 78

```

Search completed: July 7, 2004, 15:20:15  
Job time : 8.18715 secs

This Page Blank (uspto)